PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C07H 21/04, 21/02, C12N 1/21, C12P 21/00, C07K 15/28, A61K 39/395

A1

(11) International Publication Number: WO 94/12520

(43) International Publication Date: 9 June 1994 (09.06.94)

(21) International Application Number: PCT/US93/11138

(22) International Filing Date: 17 November 1993 (17.11.93)

(30) Priority Data:
07/980,529 20 November 1992 (20.11.92) US
08/002,845 15 January 1993 (15.01.93) US

(71) Applicant: ENZON, INC. [US/US]; 40 Kingsbridge Road, Piscataway, NJ 08854-3998 (US).

(72) Inventors: WHITLOW, Marc, D.; 24 Windham Court, Princeton, NJ 08540 (US). FIPULA, David; 12 Carlton Club Drive, Piscataway, NJ 08854 (US).

(74) Agents: GOLDSTEIN, Jorge, A. et al.; Sterne, Kessler, Goldstein & Fox, 1100 New York Avenue, N.W., Suite 600, Washington, DC 20005-3934 (US).

(81) Designated States: AU, BR, CA, CZ, FI, HU, JP, KR, NO, NZ, PL, RO, RU, SK, UA, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: LINKER FOR LINKED FUSION POLYPEPTIDES

(57) Abstract

The invention is directed to a novel peptide linker useful for connecting polypeptide constituents into a novel linked fusion polypeptide. The peptide linker of the invention provides greater stability and is less susceptible to aggregation than previously known peptide linkers. The peptide linker of the invention may be up to about 50 amino acids in length and contains at least one occurrence of a charged amino acid followed by a proline. When used for making a single chain Fv(sFv), the peptide linker is preferably from 18 to about 30 amino acids in length. A preferred embodiment of the peptide linker of the invention comprises the sequence: GSTSGSGXPGSGEGSTKG (SEQ ID NO 1), where X is a charged amino acid, preferably lysine or arginine. Methods of making linked fusion polypeptides using the peptide linker of the invention are claimed. DNA molecules encoding such linked fusion polypeptides, and methods of producing such linked fusion polypeptides from these DNA molecules are also claimed.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

Title of the Invention

Linker for Linked Fusion Polypeptides

Cross-Reference to Related Applications

This application is a continuation-in-part of U.S. Patent Application Serial Number 07/980,529, filed November 20, 1992.

Background of the Invention

Field of the Invention

The present invention relates to linked fusion polypeptides derived from single and multiple chain proteins. In particular, the invention relates to the linker peptide essential for bridging the polypeptide constituents that comprise the linked fusion polypeptide.

Description of the Background Art

The advent of modern molecular biology and immunology has brought about the possibility of producing large quantities of biologically active materials in highly reproducible form and with low cost. Briefly, the gene sequence coding for a desired natural protein is isolated, replicated (cloned) and introduced into a foreign host such as a bacterium, a yeast (or other fungi) or a mammalian cell line in culture, with appropriate regulatory control signals. When the signals are activated, the gene is transcribed and translated, and expresses the desired protein. In this manner, such useful biologically active materials as hormones, enzymes and antibodies have been cloned and expressed in foreign hosts.

One of the problems with this approach is that it is limited by the "one gene, one polypeptide chain" principle of molecular biology. In other words, a genetic sequence codes for a single polypeptide chain. Many biologically active polypeptides, however, are aggregates of two or more chains. For example, antibodies are three-dimensional aggregates of two heavy and two light chains. In the same manner, large enzymes such as aspartate transcarbamylase, for example, are aggregates of six catalytic and six regulatory chains, these chains being different. In order to produce such complex materials by recombinant DNA technology in foreign hosts, it becomes necessary to clone and express a gene coding for each one of the different kinds of polypeptide chains. These genes can be expressed in separate hosts. The resulting polypeptide chains from each host would then have to be reaggregated and allowed to refold together in solution. Alternatively, the two or more genes coding for the two or more polypeptide chains of the aggregate could be expressed in the same host simultaneously, so that refolding and reassociation into the native structure with biological activity will occur after expression. This approach, however, necessitates expression of multiple genes in a single host. Both of these approaches have proven to be inefficient.

Even if the two or more genes are expressed in the same organism it is quite difficult to get them all expressed in the required amounts.

A classical example of multigene expression to form multimeric polypeptides is the expression by recombinant DNA technology of antibodies. Antibodies are immunoglobulins typically composed of four polypeptides; two heavy chains and two light chains. Genes for heavy and light chains have been introduced into appropriate hosts and expressed, followed by reaggregation of these individual chains into functional antibody molecules (see, for example, Munro, *Nature 312*:597 (1984); Morrison, S.L., *Science 229*:1202' (1985); and Oi *et al., BioTechniques 4*:214 (1986); Wood *et al.*, *Nature 314*:446-449 (1985)).

Antibody molecules have two generally recognized regions in each of the heavy and light chains. These regions are the so-called "variable" region which is responsible for binding to the specific antigen in question, and the so-called "constant" region which is responsible for biological effector responses such as complement binding, etc. The constant regions are not necessary for antigen binding. The constant regions have been separated from the antibody molecule, and biologically active (i.e., binding) variable regions have been obtained.

The variable regions of a light chain (V_L) and a heavy chain (V_H) together form the structure responsible for an antibody's binding capability. Light and heavy chain variable regions have been cloned and expressed in foreign hosts, and maintain their binding ability (Moore *et al.*, European Patent Publication 0088994 (published September 21, 1983) see also Cabilly, U.S. Patent No. 4,816,567 (issued March 28, 1989)). Antibodies may be cleaved to form fragments, some of which retain their binding ability. One such fragment is the "Fv" fragment, which is composed of the terminal binding portions of the antibodies. The Fv comprises two complementary subunits, the V_L and V_H, which in the native antibody compose the binding domains.

The Fv fragment of an antibody is probably the minimal structural component which retains the binding characteristics of the parent antibody. The limited stability at low protein concentrations of the Fv fragments may be overcome by using an artificial peptide linker to join the variable domains of an Fv. The resulting single-chain Fv (hereinafter "sFv") polypeptides have been shown to have binding affinities equivalent to the monoclonal antibodies (MAbs) from which they were derived (Bird *et al., Science 242*:423 (1988)). In addition, catalytic MAbs may be converted to a sFv form with retention of catalytic characteristics (Gibbs *et al., Proc. Natl. Acad. Sci., USA 88*:4001 (1991)).

There are a number of differences between single-chain Fv (sFv) polypeptides and whole antibodies or antibody fragments, such as Fab or

$F(ab)'_2$. Single-chain Fv polypeptides are small proteins with a molecular weight around 27 kd, which lack the constant regions of 50 kd Fab fragments or 150 kd immunoglobulin antibodies bearing gamma chains (IgG). Like a Fab fragment, and unlike an IgG, an sFv polypeptide contains a single binding site.

The *in vivo* properties of sFv polypeptides are different from MAbs and antibody fragments. Due to their small size, sFv polypeptides clear more rapidly from the blood and penetrate more rapidly into tissues (Colcher, *et al.*, *J. Natl. Cancer Inst.* *82*:1191 (1990); Yokota *et al.*, *Cancer Research* *52*:3402 (1992)). Due to lack of constant regions, sFv polypeptides are not retained in tissues such as the liver and kidneys. Due to the rapid clearance and lack of constant regions, sFv polypeptides will have low immunogenicity. Thus, sFv polypeptides have applications in cancer diagnosis and therapy, where rapid tissue penetration and clearance are advantageous.

Monoclonal antibodies have long been envisioned as magic bullets, in which they deliver to a specific tumor cell a cytotoxic agent in a highly targeted manner. sFv polypeptides can be engineered with the two variable regions derived from a MAb. The sFv is formed by ligating the component variable domain genes with an oligonucleotide that encodes an appropriately designed linker polypeptide. Typically, the linker bridges the C-terminus of the first V region and the N-terminus of the second V region. sFv polypeptides offer a clear advantage over MAbs because they do not have the constant regions derived from their biological source, which may cause antigenic reaction against the MAb. Single-chain immunotoxins have been produced by fusing a cell binding sFv with *Pseudomonas* exotoxin (Chaudhary *et al.*, *Nature* *339*:394 (1989)). Recently, a single-chain immunotoxin was shown to cause tumor regression in mice (Brinkmann *et al.*, *Proc. Natl. Acad. Sci. USA* *88*:8616 (1991)).

The general considerations behind the design and construction of polypeptide linkers as applied to sFv polypeptides have been previously described in U.S. Patent No. 4,946,778 (Ladner *et al.*). Computer design of

linkers has also been described in U.S. Patent Nos. 4,704,692, 4,853,871, 4,908,773 and 4,936,666.

Four linkers are described in the '778 disclosure: TRY40, TRY 59, TRY61, and TRY104b. TRY40 is a double linker with 3- and 7-amino acid sequences comprising the linkers. The sequences are PGS and IAKAFKN (see page 8, Table 1 for a description of the single letter amino acid code used herein). TRY59 is an 18-residue single linker having the sequence KESGSVSSEQLAQFRSLD (SEQ. ID No. 2). TRY 61 is a 14-residue single linker having the sequence VRGSPAINVAVHVF (SEQ. ID No. 3). TRY104b is a 22-residue single linker constructed primarily of a helical segment from human hemoglobin. The sequence is AQGTLSPADKTNV KAAWGKVMT (SEQ. ID No. 4).

Traunecker *et al.*, *EMBO J. 10(12)*:3655-3659 (1991) have disclosed an 18-amino acid linker for joining the first two N-terminal CD4 domains and the combining site of the human CD3 complex. Its sequence is VEGGSGGS GGSGGSGGVD (SEQ. ID No. 5). The final bispecific single-chain polypeptide is called Janusin, and targets cytotoxic lymphocytes on HIV-infected cells.

Fuchs *et al.*, *Bio/Technology 9*:1369-1372 (1991), used an 18-residue linker to join the heavy- and light-chain variable domains of a humanized antibody against chick lysozyme. The 18-residue linker was partially derived from α-tubulin and contains a MAb epitope specific to α-tubulin. The full sequence is GSASAPKLEEGEFSEARE (SEQ. ID No. 6).

A host of single-chain Fv analog polypeptides are disclosed in the literature (see, Huston, J.S. *et al.*, *Proc. Natl. Acad. Sci. USA 85*:5879-5883 (1988); Huston, J.S. *et al.*, *SIM News 38(4) (Suppl.)*:11 (1988); McCartney, J. *et al.*, *ICSU Short Reports 10*:114 (1990); McCartney, J.E. *et al.*, unpublished results (1990); Nedelman, M.A. *et al.*, *J. Nuclear Med. 32 (Suppl.)*:1005 (1991); Huston, J.S. *et al.*, In: *Molecular Design and Modeling: Concepts and Applications, Part B*, edited by J.J. Langone, *Methods in Enzymology 203*:46-88 (1991); Huston, J.S. *et al.*, In: *Advances*

in the Applications of Monoclonal Antibodies in Clinical Oncology, Epenetos, A.A. (Ed.), London, Chapman & Hall (in preparation 1992); Bird, R.E. *et al.*, *Science 242*:423-426 (1988); Bedzyk, W.D. *et al.*, *J. Biol. Chem. 265*:18615-18620 (1990); Colcher, D. *et al.*, *J. Nat. Cancer Inst. 82*:1191-1197 (1990); Gibbs, R.A. *et al.*, *Proc. Natl. Acad. Sci. USA 88*:4001-4004 (1991); Milenic, D.E. *et al.*, *Cancer Research 51*:6363-6371 (1991); Pantoliano, M.W. *et al.*, *Biochemistry 30*:10117-10125 (1991); Chaudhary, V.K. *et al.*, *Nature 339*:394-397 (1989); Chaudhary, V.K. *et al.*, *Proc. Natl. Acad. Sci. USA 87*:1066-1070 (1990); Batra, J.K. *et al.*, *Biochem. Biophys. Res. Comm. 171*:1-6 (1990); Batra, J.K. *et al.*, *J. Biol. Chem. 265*:15198-15202 (1990); Chaudhary, V.K. *et al.*, *Proc. Natl. Acad. Sci. USA 87*:9491-9494 (1990); Batra, J.K. *et al.*, *Mol. Cell. Biol. 11*:2200-2205 (1991); Brinkmann, U. *et al.*, *Proc. Natl. Acad. Sci. USA 88*:8616-8620 (1991); Seetharam, S. *et al.*, *J. Biol. Chem. 266*:17376-17381 (1991); Brinkmann, U. *et al.*, *Proc. Natl. Acad. Sci. USA 89*:3075-3079 (1992); Glockshuber, R. *et al.*, *Biochemistry 29*:1362-1367 (1990); Skerra, A. *et al.*, *Bio/Technol. 9*:273-278 (1991); Pack, P. *et al.*, *Biochem. 31*:1579-1534 (1992); Clackson, T. *et al.*, *Nature 352*:624-628 (1991); Clackson, T. *et al.*, *Nature 352*:624-628 (1991); Marks, J.D. *et al.*, *J. Mol. Biol. 222*:581-597 (1991); Iverson, B.L. *et al.*, *Science 249*:659-662 (1990); Roberts, V.A. *et al.*, *Proc. Natl. Acad. Sci. USA 87*:6654-6658 (1990); Condra, J.H. *et al.*, *J. Biol. Chem. 265*:2292-2295 (1990); Laroche, Y. *et al.*, *J. Biol. Chem. 266*:16343-16349 (1991); Holvoet, P. *et al.*, *J. Biol. Chem. 266*:19717-19724 (1991); Anand, N.N. *et al.*, *J. Biol. Chem. 266*:21874-21879 (1991); Fuchs, P. *et al.*, *Bio/Technol. 9*:1369-1372 (1991); Breitling, F. *et al.*, *Gene 104*:104-153 (1991); Seehaus, T. *et al.*, *Gene 114*: in press (1992); Takkinen, K. *et al.*, *Prot. Eng. 4*:837-841 (1991); Dreher, M.L. *et al.*, *J. Immunol. Methods 139*:197-205 (1991); Mottez, E. *et al.*, *Eur. J. Imunol. 21*:467-471 (1991); Traunecker, A. *et al.*, *Proc. Natl. Acad. Sci. USA 88*:8646-8650 (1991); Traunecker, A. *et al.*, *EMBO J. 10*:3655-3659 (1991); Hoo, W.F.S. *et al.*, *Proc. Natl. Acad. Sci.*

USA 89:4759-4763 (1993)). Linker lengths used in those Fv analog polypeptides vary from 10 to 28 residues.

Linkers previously used for sFvs and other polypeptides suffer from proteolytic attack, rendering them less stable and prone to dissociation. They also suffer from inordinate aggregation at high concentrations, making them susceptible to concentration in the liver and kidneys. Therefore, there is a need for more stable linkers that are resistant to proteolytic attack and less prone to aggregation.

Summary of the Invention

The invention is directed to a linked fusion polypeptide comprising polypeptide constituents connected by a novel peptide linker. The novel peptide linker comprises a sequence of amino acids numbering from about 2 to about 50 having a first end connected to a first protein domain, and having a second end connected to a second protein domain, wherein the peptide comprises at least one proline residue within the sequence, the proline being positioned next to a charged amino acid, and the charged amino acid-proline pair is positioned within the peptide linker to inhibit proteolysis of said polypeptide.

The invention is also directed to a novel peptide linker comprising the amino acid sequence:

GSTSGSGXPGSGEGSTKG (SEQ ID NO 1),

wherein the numbering order from left to right (amino to carboxyl) is 1 to 18, and X is a charged amino acid. In a preferred embodiment X is lysine or arginine.

The invention also relates to sFvs wherein the linker linking V_H and V_L regions is the peptide linker as herein described, preferably comprising from about 10 to about 30 amino acids, and more preferably comprising at least 18 amino acids.

The invention also relates to genetic sequences encoding linked fusion polypeptides containing the novel peptide linker herein described, methods of making such linked fusion polypeptides, and methods of producing such linked fusion polypeptides via recombinant DNA technology.

Brief Description of the Drawings

Figure 1 is a set of two graphs depicting the proteolytic susceptibility of the CC49/212 and CC49/218 sFv proteins when exposed to *subtilisin BPN'* (Panel A) or trypsin (Panel B). The fraction of sFv remaining intact was determined by reverse phase HPLC. The CC49/212 sFv is shown in open circles and the CC49/218 is shown in closed squares. There was no measurable degradation of the CC49/218 sFv.

Figure 2 is a graph depicting the results of a competition radioimmunoassay (RIA) in which unlabeled CC49/212 single-chain Fv (open squares), CC49/218 single-chain Fv (closed diamonds) or MOPC-21 IgG (+) competed against a CC49 IgG radiolabeled with ^{125}I for binding to the TAG-72 antigen on a human breast carcinoma extract. MOPC-21 is a control antibody that does not bind to TAG-72 antigen.

Figure 3 is the amino acid (SEQ. ID No. 12) and nucleotide (SEQ. ID No. 11) sequence of the linked fusion polypeptide comprising the 4-4-20 V_L region connected through the 217 linker to the CC49 V_H region.

Figure 4 is the amino acid (SEQ. ID No. 14) and nucleotide (SEQ. ID No. 13) sequence of the linked fusion polypeptide comprising the CC49 V_L region connected through the 217 linker polypeptide to the 4-4-20 V_H region.

Figure 5 is a chromatogram depicting the purification of CC49/4-4-20 heterodimer Fv on a cation exchange high performance liquid chromatographic column. The column is a PolyCAT A aspartic acid column (Poly LC, Columbia, MD). The heterodimer Fv is shown as peak 5, eluting at 30.10 min.

Figure 6 is a coomassie-blue stained 4-20% SDS-PAGE gel showing the proteins separated in Figure 5. Lane 1 contains the molecular weight standards. Lane 3 contains the starting material before separation. Lanes 4-8 contain fractions 2, 3, 5, 6 and 7, respectively. Lane 9 contains purified CC49/212.

Figure 7 is a chromatogram used to determine the molecular size of fraction 2 from Figure 5. A TSK G3000SW gel filtration HPLC column was used (Toyo Soda, Tokyo, Japan).

Figure 8 is a chromatogram used to determine the molecular size of fraction 5 from Figure 5. A TSK G3000SW gel filtration HPLC column was used (Toyo Soda, Tokyo, Japan).

Figure 9 is a chromatogram used to determine the molecular size of fraction 6 from Figure 5. A TSK G3000SW gel filtration HPLC column was used (Toyo Soda, Tokyo, Japan).

Figure 10 shows a Scatchard analysis of the fluorescein binding affinity of the CC49/4-4-20 heterodimer Fv (fraction 5 in Figure 5).

Figure 11 is a graphical representation of three competition enzyme-linked immunosorbent assays (ELISA) in which unlabeled CC49/4-4-20 Fv (closed squares) CC49/212 single-chain Fv (open squares) and MOPC-21 IgG (+) competed against a biotin-labeled CC49 IgG for binding to the TAG-72 antigen on a human breast carcinoma extract. MOPC-21 is a control antibody that does not bind to the TAG-72 antigen.

Definitions

Amino acid Codes

The most common amino acids and their codes are described in Table 1:

Table 1

Amino acid names and codes

Amino acid	Single letter code
Alanine	A
Arginine	R
Aspartic acid	D
Asparagine	N
Cysteine	C
Glutamic acid	E
Glutamine	Q
Glycine	G
Histidine	H
Isoleucine	I
Leucine	L
Lysine	K
Methionine	M
Phenylalanine	F
Proline	P
Serine	S
Threonine	T
Tryptophane	W
Tyrosine	Y
Valine	V

Protein: As referred to herein, a protein is a biological molecule which consists primarily of one or more polypeptides. A protein consisting of a single polypeptide is referred to herein as a **single chain protein**. A protein consisting of more than one polypeptide is referred to herein as a **multi-chain protein**, with the term chain being synonymous with the term polypeptide.

Polypeptide: As referred to herein, a polypeptide is a linear, single chain polymer of multiple amino acids linked through their amino and carboxylate groups by peptide bonds. A polypeptide may form a single chain protein by itself or, in association with other polypeptides, form a multi-chain protein. A polypeptide may also be a fragment of a single chain protein or a fragment of one of the chains of a multi-chain protein.

Linked fusion polypeptide: As referred to herein, a linked fusion polypeptide is a polypeptide made up of two smaller polypeptide constituents, each constituent being derived from a single chain protein or a single chain of a multi-chain protein, where the constituents are combined in a non-naturally occurring arrangement using a peptide linker. Linked fusion polypeptides mimic some or all of the functional aspects or biological activities of the protein(s) from which their polypeptide constituents are derived. The constituent at the amino terminal portion of the linked fusion polypeptide is referred to herein as the first polypeptide. The constituent at the carboxy terminal portion of the linked fusion polypeptide is referred to herein as the second polypeptide. By "non-naturally occurring arrangement" is meant an arrangement which occurs only through *in vitro* manipulation of either the polypeptide constituents themselves or the nucleic acids which encode them.

Peptide linker: As referred to herein, a peptide linker or linker is a polypeptide typically ranging from about 2 to about 50 amino acids in length, which is designed to facilitate the functional connection of two polypeptides into a linked fusion polypeptide. The term functional connection denotes a connection that facilitates proper folding of the polypeptides into a three dimensional structure that allows the linked fusion polypeptide to mimic some or all of the functional aspects or biological activities of the protein(s) from which its polypeptide constituents are derived. In cases such as sFv polypeptides where the linker is used to make a single chain derivative of a multi-chain protein, to achieve the desired biological activity the appropriate

three dimensional structure will be one that mimics the structural relationship of the two polypeptide constituents in the native multi-chain protein. The term functional connection also denotes a connection that confers a degree of stability required for the resulting linked fusion polypeptide to function as desired.

Charged Amino Acid: As referred to herein, a charged amino acid is a biologically derived amino acid which contains a charge at neutral pH. Charged amino acids include the negatively charged amino acids Aspartic acid (D) and Glutamic acid (E) as well as positively charged amino acids Histidine (H), Lysine (K), and Arginine (R).

Immunoglobulin superfamily: As referred to herein, the immunoglobulin superfamily is the family of proteins containing one or more regions that resemble the variable or constant regions of an immunoglobulin, or fundamental structural units (i.e., domains) found within these regions. The resemblance referred to is in terms of size, amino acid sequence, and presumably three dimensional structure. Members of the immunoglobulin superfamily typically mediate non-enzymatic intercellular surface recognition and include, but are not limited to, CD1, CD2, CD3, CD7, CD8, CD28 class I and II histocompatibility molecules, Beta-2 microglobulin, lymphocyte function associated antigen-3 (LFA-3), Fc_γ receptor, Thy-1, T cell receptor, polyimmunoglobulin receptor, neuronal cell adhesion molecule, myelin associated glycoprotein, P_o myelin, carcinoembryonic antigen, platelet derived growth factor receptor, colony stimulating factor-1 receptor, link protein of basement membrane, and $\alpha_1\beta$-glycoprotein.

T cell Receptor: As referred to herein, T cell receptor is a member of the immunoglobulin superfamily that resides on the surface of T lymphocytes and specifically recognizes molecules of the major histocompatibility complex, either alone or in association with foreign antigens.

Immunoglobulin: As referred to herein, an immunoglobulin is a multi-chain protein with antibody activity typically composed of two types of polypeptides, referred to as heavy and light chains. The heavy chain is larger than the light chain and typically consists of a single variable region, three or four constant regions, a carboxy-terminal segment and, in some cases, a hinge region. The light chain typically consists of a single variable region and a single constant region.

Antibody: As referred top herein, an antibody is an immunoglobulin that is produced in response to stimulation by an antigen and that reacts specifically with that antigen. Antibodies are typically composed of two identical heavy and two identical light polypeptide chains, held together by interchain disulfide bonds.

Single chain Fv polypeptide (sFv): As referred to herein, a single chain Fv polypeptide (sFv) is a linked fusion polypeptide composed of two variable regions derived from the same antibody, connected by a peptide linker. An sFv is capable of binding antigen similar to the antibody from which its variable regions are derived. An sFv composed of variable regions from two different antibodies is referred to herein as a mixed sFv.

Detailed Description of the Invention

In order to design a peptide linker that will join any multichain protein to form a linked fusion polypeptide with the same or similar function as the multi-chain protein, it is necessary to define the extent of each chain that must be included. For example, to design a peptide linker that will join the variable domains of an antibody to form an sFv, the extent of the variable domains must first be defined. Kabat *et al.* (Kabat *et al.*, Sequences of Proteins of Immunological Interest, Department of Health and Human Services, Fourth Edition, U.S. (1987)) defines the variable domain (V_L) to extend from residue

1 to residue 107 for the lambda light chain, and to residue 108 for kappa light chains, and the variable domain of the heavy chain (V_H) to extend from residue 1 to residue 113.

Single-chain Fvs can and have been constructed in several ways. Either V_L is the N-terminal domain followed by the linker and V_H (a V_L-Linker-V_H construction) or V_H is the N-terminal domain followed by the linker and V_L (V_H-Linker-V_L construction). Alternatively, multiple linkers have also been used. Several types of sFv proteins have been successfully constructed and purified, and have shown binding affinities and specificities similar to the antibodies from which they were derived.

Typically, the Fv domains have been selected from the group of monoclonal antibodies known by their abbreviations in the literature as 26-10, MOPC 315, 741F8, 520C9, McPC 603, D1.3, murine phOx, human phOx, RFL3.8 sTCR, 1A6, Se155-4, 18-2-3, 4-4-20, 7A4-1, B6.2, CC 49, 3C2, 2c, MA-15C5/$K_{12}G_0$, Ox, etc. (see references previously cited as disclosing Fv analog polypeptides). One of ordinary skill in the art will be able to adapt a linker to join other domains not mentioned herein. The Fv's are derived from the variable regions of the corresponding monoclonal antibodies (MAbs).

Linkers have also been used to join non-antibody polypeptides, as evidenced by Soo Hoo *et al.*, *Proc. Natl. Acad. Sci. USA 89*:4759-4763 (1992) and Kim *et al. Protein Engineering 2(8)*:571-575 (1989). Soo Hoo *et al.* discloses a linker connecting the variable regions of the α and β chains of a T cell receptor. Kim *et al.* discloses a linker designed to link the two polypeptide chains of monellin, a multi-chain protein known for its sweet taste.

Thus, it is envisioned that linkers according to the invention will be useful for connecting polypeptides derived from any protein. The order in which the polypeptides are connected (i.e., which is nearer the amino or carboxy terminus of the linked fusion polypeptide) should, where possible, reflect the relationship of the polypeptides in their native state. For example, consider a linked fusion polypeptide derived from two chains of a multi-chain protein where the amino terminal portion of the first chain is normally

associated (i.e., in proximity to) the carboxy terminal portion of the second chain. In this case, the polypeptide derived from the first chain should be positioned near the amino-terminal portion of the linked fusion polypeptide and the polypeptide derived from the second chain should be positioned near the carboxy-terminal portion.

In particular, it is envisioned that linkers according to the invention will be applicable to any multi-chain protein or protein complex including, but not limited to, members of the immunoglobulin superfamily, enzymes, enzyme complexes, ligands, regulatory proteins, DNA-binding proteins, receptors, hormones, etc. Specific examples of such proteins or protein complexes include, but are not limited to, T cell receptors, insulin, RNA polymerase, Myc, Jun, Fos, glucocorticoid receptor, thyroid hormone receptor, acetylcholine receptor, fatty acid synthetase complex, hemoglobin, tubulin, myosin, β-Lactoglobulin, aspartate transcarbamoylase, malic dehydrogenase, glutamine synthetase, hexokinase, glyceraldehyde-phosphate dehydrogenase, glycogen phosphorylase, tryptophan synthetase, etc.

It is also envisioned that non-polypeptide biochemical moieties including, but not limited to, toxins, drugs, radioisotopes, etc. may be added to, or associated with, the linked fusion polypeptides to achieve a desired effect, such as labeling or conferring toxicity.

The preferred length of the peptide linker should be from 2 to about 50 amino acids. In each particular case, the preferred length will depend upon the nature of the polypeptides to be linked and the desired activity of the linked fusion polypeptide resulting from the linkage. Generally, the linker should be long enough to allow the resulting linked fusion polypeptide to properly fold into a conformation providing the desired biological activity. Where conformational information is available, as is the case with sFv polypeptides discussed below, the appropriate linker length may be estimated by consideration of the 3-dimensional conformation of the substituent polypeptides and the desired conformation of the resulting linked fusion polypeptide. Where such information is not available, the appropriate linker

length may be empirically determined by testing a series of linked fusion polypeptides with linkers of varying lengths for the desired biological activity.

Linkers of the invention used to construct sFv polypeptides are designed to span the C terminus of V_L (or neighboring site thereof) and the N terminus of V_H (or neighboring site thereof) or between the C terminus of V_H and the N terminus of V_L. The linkers used to construct sFv polypeptides have between 10 and 30 amino acid residues. The linkers are designed to be flexible, and it is recommended that an underlying sequence of alternating Gly and Ser residues be used.

To enhance the solubility of the linker and its associated single chain Fv protein, three charged residues may be included, two positively charged lysine residues (K) and one negatively charged glutamic acid residue (E). Preferably, one of the lysine residues is placed close to the N-terminus of V_H, to replace the positive charge lost when forming the peptide bond of the linker and the V_H.

In addition, it has unexpectedly been found that linker lengths of equal to or greater than 18 residues reduce aggregation. This becomes important at high concentrations, when aggregation tends to become evident. Thus, linkers having 18 to 30 residues are preferred for sFv polypeptides.

Another property that is important in engineering an sFv polypeptide, or any other linked fusion polypeptide, is proteolytic stability. The 212 linker (Pantoliano *et al.*, *Biochemistry 30*:10117 (1991)) is susceptible to proteolysis by subtilisin BPN'. The proteolytic clip in the 212 linker occurs between Lys8 and Ser9 of the linker (see Table 2). By placing a proline at the proteolytic clip site one may be able to protect the linker. The inventors, not wishing to be bound by any particular theory of operation, postulate that the proline residue in the peptide linker of the present invention inhibits the charge-transfer intermediate that is essential to the hydrolysis of the amide bond joining the two amino acid residues clipped apart by serine proteases.

Table 2 shows two of the claimed linkers (217 and 218) and two of the prior art linkers (202' and 212) for illustration. The 217 linker contains a

lysine-proline pair at positions 6 and 7, thus rendering the linker less susceptible to proteolysis. The 218 linker demonstrates less aggregation, proteolytic stability, and the necessary flexibility and solubility to result in a functional linker for sFv proteins.

Table 2

Linker Designs

V_L-Linker-V_H Construction

V_L	Linker	V_H	Linker Name	Reference
-KLEIE	GKSSGSGSESKS[3]	TQKLD-	202′	Bird *et al.*[1]
-KLEIK	GSTSGSGKSSEGKG[4]	EVKLD-	212	Bedzyk *et al.*[2]
-KLEIK	GSTSGSGKSSEGSGSTKG[5]	EVKLD-	216	212 Experimental Derivative
-KLVLK	GSTSGKPSEGKG[6]	EVKLD-	217	Invention
-KLEIK	GSTSGSGKPGSGEGSTKG[7]	EVKLD-	218	Invention

(1) *Science 242*:423 (1988)
(2) *JBC 265*:18615-18620 (1990)
(3) SEQ ID No. 7
(4) SEQ ID No. 8
(5) SEQ. ID No. 9
(6) Part of SEQ ID No. 12
(7) SEQ ID No. 10

The stability and affinity of an antifluorescein single-chain Fv's has been previously reported (Pantoliano, M.W., *et al.*, *Biochemistry 30*:10117-10125 (1991)). The data in the prior studies showed that the affinity of the 4-4-20 sFvs for fluorescein may increase with longer linkers. The data was not conclusive for the longest linker, 205, which was thought to be helical. Thus, a 4-4-20 sFv was designed, constructed, purified and assayed with an 18 residue linker that was four residues longer than the 212 linker (see Table 2). This new linker was designated 216. The anti-fluorescein sFvs 4-4-20/202′, 4-4-20/212 and 4-4-20/216 had affinities of 0.5 x 10^9 M^{-1}, 1.0 x 10^9 M^{-1}, and 1.3 x 10^9 M^{-1}, respectively using the fluorescence quenching assay.

In attempting to crystalize the anti-fluorescein 4-4-20 sFvs, they were concentrated to over 10 mg/ml. At these high concentrations it was noticed that they produced aggregates under a wide variety of conditions, as judged by size-exclusion HPLC chromatography. Although aggregation can be reversed by diluting the sample, it is an undesirable phenomenon. It was discovered that shorter linkers showed higher degrees of aggregation than larger linkers. For example, at 5 mg/ml the 4-4-20/202′ sFv sample was 53% aggregated, whereas the 4-4-20/212 and 4-4-20/216 samples showed 34% and 10% aggregation, respectively.

A second discovery made in trying to crystallize the anti-fluorescein 4-4-20 sFvs was that the prior art 212 linker was proteolytically susceptible. It was possible to produce crystals of the 4-4-20/212 sFv only after it had been treated with subtilisin BPN′, a serine protease. When 4-4-20/212 sFv and subtilisin *BPN′* were mixed in a 5000 to 1 ratio, the 27 kD band of the sFv was converted into two bands that ran just below the 14 kD marker on the SDS-PAGE. N-terminal sequencing of the clipped sFv showed that the prior art 212 linker had been clipped between the Lys 8 and Ser 9. The effective result of this clip was to change a sFv into an Fv, a much less stable molecule.

Without being bound to any particular theory underlying the invention, the inventors believe that the following discussion may explain the markedly improved characteristics of the 218 linker and other such linkers. In order to reduce the proteolytic susceptibility of the sFvs it is possible to protect the susceptible peptide bond between Lys 8 and Ser 9 in the linker of the invention. Most proteases are unable to cleave peptide-located bonds prior to a proline. This is because prolines do not have amide hydrogens. The proline side chain forms a five-membered ring with the amide nitrogen. It is believed that the five-membered ring of the proline prohibits proteolysis from occurring. It is believed that proline is unique in its ability to so inhibit proteolysis. Placement of the proline next to a charged residue is also preferred. The sequence of proline and a charged amino acid residue should be maintained with the charged residue before (i.e., on the amino-terminus

side of) the proline. In a preferred embodiment, a lysine-proline pair is located at the cleavage site, replacing the susceptible amide bond that was hydrolyzed. In a second preferred embodiment, arginine may be used as the charged residue.

A second guiding consideration in designing the linker of the invention is that a linker with reduced aggregation is preferable. As described above, the 18-residue 216 linker shows reduced aggregation as compared to the 14-residue 212 linker. The first eleven residues of the 216 linker are identical to the 212 linker, including the proteolytically-susceptible peptide bond between Lys 8 and Ser 9. Thus, it is believed that the extra four residues contribute to the lowered aggregation. Linkers with 18 or more residues are thus preferred.

Taking the above into consideration, a new linker was designed with a Lys-Pro sequence at positions 8 and 9 and a length of 18 amino acids. This linker was then subjected to testing in order to prove that it has the characteristics it was designed to have. The new linker was designated 218 (see Table 2).

Positioning the proline at the proper place in the linker sequence to inhibit proteolysis is accomplished by determining the points of proteolytic attack in the susceptible sequence. One of ordinary skill in the art will know of methods of determining this point. In one method, a protease such as subtilisin BPN′ is contacted with the candidate linker. Cleavage can then be determined by sequencing the resulting peptides, which will also reveal the cleavage point or points, if any. Any protease may be used, and selection will be guided by consideration of the environment the linker is to encounter in actual use.

Also provided by the invention are DNA molecules such as isolated genetic sequences or plasmids or vectors encoding linked fusion polypeptides with the peptide linker of the invention. The DNA sequence for the linked fusion polypeptide can be chosen so as to optimize production in organisms such as bacteria or yeast.

Recombinant hosts as well as methods of using them to produce single chain proteins by expression, are also provided herein.

The appropriate DNA molecules, hosts, methods of production, isolation and purification of linked fusion polypeptides, especially sFv polypeptides, are thoroughly described in the prior art, such as e.g., U.S. Patent No. 4,946,778, which is fully incorporated herein by reference.

Examples

1. General Test Conditions

Cloning and Genetic Constructions. The cloning of the 4-4-20 variable domains has been previously described by Bedzyk, W.D., *et al.*, *J. Biol. Chem. 264*:1565-1569 (1989). The sequence of the variable domain of the CC49 domain has been previously described by Mezes, P., *et al.*, European Patent Application No. EP 0 365 997 (1989). The genetic construction of the 4-4-20/202', 4-4-20/212 and CC49/212 sFvs have been previously described by Bedzyk, W.D., *et al.*, *J. Biol. Chem. 265*:18615-18620 (1990) or Pantoliano, M.W., *et al.*, *Biochemistry 30*:10117-10125 (1991) and Milenic, D., *et al.*, *Cancer Res. 51*:6363-6371 (1991), respectively.

Purification. The purification of sFv polypeptides has been previously described by Pantoliano, M.W., *et al.*, *Biochemistry 30*:10117-10125 (1991) and Whitlow and Filpula, *Methods 2*:97-105 (1991). Most of the sFv polypeptides were purified with a minor procedural modification, omitting the initial cation exchange HPLC step using the RCM Waters Accell Plus GM ion exchange (RCM) column.

Association constants of the anti-fluorescein sFvs. The association constants were determined for each of the anti-fluorescein sFvs following the procedures described by Herron and Voss, in *Fluorescence Hapten: An*

Immunological Probe, E.W. Voss, Jr., ed., CRC Press, Boca Raton, FL, 77-98 (1984).

Aggregation Rates. The rates of aggregation of the sFv polypeptides were determined at room temperature in 60 mM MOPS, pH 7.0 at various concentrations using Gel Filtration HPLC Chromatography. 10 to 50 μl samples were injected onto a Waters HPLC system with 7.8 mm x 300 mm TSK G3000SW column (Toso Haas, Tokyo, Japan). The column had been previously equilibrated and the samples were eluted using 50 mM MOPS, 100 mM NaCl, buffer pH 7.5 at a flow rate of 0.5 ml/min. The data was collected on a MacIntosh SE (Apple Computer, Cupertino, CA) running the Dynamac software package (Rainin Instrument Co, Woburn, MA).

Radiolabeling of Proteins. MAb CC49 and CC49 sFv polypeptides were labeled with $Na^{125}I$ using Iodo-Gen (Pierce Chemical Co., Rockford, IL) as previously reported (Milenic, D., *et al.*, *Cancer Res.* *51*:6363-6371 (1991)).

The CC49 sFv polypeptides were labeled with the lutetium complex of the macrocyclic bifunctional coordinator PA-DOTA (Cheng *et al.*, European Patent Application No. 353,450). 20 μl of a 1 mM solution of SCN-PA-DOTA in water was mixed with equal volumes of the $^{177}Lu(NO_3)_3$ solution and 1 M HEPES buffer pH 7.0 and left at room temperature for five minutes. ^{177}Lu in 0.05 N HCl was obtained from the University of Missouri Research Reactor (Columbia, MO). The reaction mixture was processed over a PRP-1 reverse-phase cartridge (Hamilton Co., Reno, NV) which had been equilibrated with 10% acetonitrile in 20 mM sodium carbonate, pH 9.5. ^{177}Lu-SCN-PA-DOTA was eluted with acetonitrile/carbonate buffer (1:2) and a 60 μl fraction containing the radioactive chelate was used.

1 mg of each CC49 sFv was exchanged with 20 mM sodium carbonate, pH 9.5 buffer, then made to 980 μl with the same buffer. The sample was mixed with 20 ml of the ^{177}Lu-SCN-PA-DOTA solution and left for 3 hours at 37°C, followed by PD-10 isolation as above. Both radiolabeling procedures resulted in >90% acid-precipitable counts.

2. Proteolytic Susceptibility of the 218 Linker

$1.0 \pm 0.1 \times 10^{-5}$ M CC49/212 and CC49/218 sFv polypeptides were treated either with 2.6×10^{-7} M subtilisin BPN' (Type XXVII protease, Sigma, St. Louis, MO) or with 7.7×10^{-7} M trypsin at 37°C. The percent sFv remaining was monitored by reverse phase HPLC at various times. A non-linear gradient between 5% acetonitrile, 0.1% TFA and 70% acetonitrile, 0.1% TFA was run on a PLRP-S column (Polymer Labs., Church Stretton, England) in a heating unit (Timberline Instruments, Boulder, CO) on a waters HPLC system, following the procedures of Nugent, K.D., *Am. Biotechnol. Lab.*, pp. 24-32 (May 1990). The data was collected on a MacIntosh SE (Apple Computer, Cupertino, CA) running the Dynamac software package (Rainin Instrument Co, Woburn, MA). The half-life ($t_{1/2}$) was determined from plots of the log of the fraction of sFv remaining versus time (Figure 1).

The half-life of the CC49/212 sFv treated with subtilisin or trypsin is 122.8 min and 195.7 min, respectively (see Figure 1). The 218 linker had significantly improved protease resistance, for in the 48 hour period digestion of the CC49/218 sFv was not detectable using either subtilisin or trypsin.

3. Binding Affinity with the 218 Linker

To determine the binding properties of the CC49 sFv polypeptides a competition radioimmunoassay (RIA) was set up in which a CC49 IgG labeled with ^{125}I was competed against the unlabeled CC49 sFvs for binding to TAG-72 on a human breast carcinoma extract as previously described by Milenic, D., *et al.*, *Cancer Res. 51*:6363-6371 (1991).

The binding affinities for the TAG-72 antigen of the CC49/212 and CC49/218 sFv polypeptides were checked. The CC49/218 sFv showed about a 4-fold lower affinity than the CC49/212 sFv (see Figure 2). The lower affinity of the CC49/218 sFv could be in part due to the higher degree of aggregation of the CC49/212 sFv sample. We have shown previously that the

dimeric forms of CC49 (IgG and $F(ab')_2$) compete with a ten-fold higher affinity than do the monovalent forms (Fab and sFv) (Milenic, D., *et al.*, *Cancer Res. 51*:6363-6371 (1991)). Since aggregates are multivalent it seems likely that they would have high affinity.

4. Aggregation Rates with 218 Linker

The rates of aggregation of the CC49/212 and CC49/218 sFv polypeptides were determined at room temperature (22°C) at various concentrations. The CC49/212 sFv showed 80-fold faster accumulation of aggregates than did the CC49/218 sFv, at concentrations around 1.5 mg/ml (see Table 3). At 0.5 mg/ml this difference increased to 1600-fold. The aggregation of both proteins showed a concentration dependence. The higher the concentration the higher the levels of aggregation that were seen.

5. Comparison of 212 and 218 Linkers in vivo

Both the observation that longer linkers result in less aggregation and that linkers could be proteolytically susceptible have possible implications in the *in vivo* therapeutic applications of sFv polypeptides, as well as other linked fusion polypeptides. First, aggregation could result in the unwanted accumulation of sFv in non-target tissues. Second, the proteolysis of a sFv to an Fv is likely to result in a loss of affinity. These two effects were examined *in vivo* in a human tumor model system. We examined the *in vivo* performance of the CC49/212 and CC49/218 sFvs in an LS-174T tumor xenograft in athymic nude mice.

Female athymic nude mice (nu/nu), obtained from Charles River (Wilmington, MA) at 4-6 weeks of age, were injected subcutaneously on the back with 1 x 10^6 LS-174T human colon carcinoma cells under a NIH-approved protocol (Tom, R.H., *et al.*, *In Vitro (Rockville) 12*:180-191 (1976)). Animals were used in biodistribution studies when the animals'

tumors measured 0.5 to 0.8 cm in diameter, approximately two weeks later. Dual-label studies were performed with tumor-bearing mice injected via the tail vein with approximately 2-10 x 10^6 cpm of each labeled CC49 sFv. Mice (3-4/data point) were killed at various time points by exsanguination. The blood, tumor and all the major organs were collected, wet-weighed and counted in a gamma scintillation counter. The % injected dose per gm (%ID/g) and radiolocalization index (%ID/g in the tumor divided by the %ID/g in normal tissue) for each were determined.

The biodistribution of the ^{177}Lu labeled CC49/212 and CC49/218 sFv polypeptides was determined at various times in athymic nude mice bearing the two-week old human colon carcinomas. Of the six tissues examined, three tissues showed significant differences between the CC49/212 and CC49/218 sFvs (see Table 4). The spleen and the liver showed three- to four-fold higher accumulations of the CC49/212 sFv compared to the CC49/218 sFv. At the 24 and 48 hour time points the CC49/212 sFv showed a 60% higher accumulation at the tumor. The other three tissues (blood, kidney and lung) show little or no differences.

The higher level of CC49/212 sFv accumulation in the spleen and liver is likely due to the higher degree of aggregation of the sample injected. Both the spleen and liver metabolize the sFv polypeptides, but due to the higher degree of aggregation of the CC49/212 sFv higher uptake and accumulation of the ^{177}Lu radiolabel in these tissues is seen. The higher levels of CC49/212 sFv in the tumor at later times may be due to the increased avidity of the aggregates. The very high levels of accumulation of both sFv polypeptides in the kidneys probably reflects the catabolism of the protein in the kidneys, with subsequent retention of the ^{177}Lu (Schott *et al.*, submitted).

Table 3

Aggregation Rates of the CC49/212 and CC49/218 sFvs

Protein	Concentration (mg/ml)	Rate of Aggregation (%/hr)	Rate of Aggregation (%/day)
CC49/212	1.89 0.49	0.732 0.120	17.56 2.88
CC49/218	1.49 0.62	0.0092 0.00008	0.221 0.0018

Table 4

Biodistribution of the ^{177}Lu labeled CC49/212 and CC49/218 sFvs

		% ID / gm			
Organ	Liver	1 h	6 h	24 h	48 h
Tumor	212	2.4	2.0	2.2	1.6
	218	2.6	1.9	1.4	1.0
	212/218 ratio	0.9	1.0	1.6	1.6
Blood	212	1.8	0.2	<0.1	<0.1
	218	0.9	0.2	<0.1	<0.1
	212/218 ratio	2.0	1.0	–	–
Liver	212	7.4	9.4	5.5	4.0
	218	3.1	2.3	1.8	1.1
	212/218 ratio	2.4	4.1	3.1	3.6
Spleen	212	9.6	7.0	7.2	6.8
	218	3.1	2.1	1.9	1.6
	212/218 ratio	3.1	3.3	3.8	4.2
Kidney	212	241.1	219.1	197.6	156.1
	218	303.9	266.0	222.9	161.5
	212/218 ratio	0.8	0.8	0.9	1.0
Lung	212	1.7	0.8	0.7	0.5
	218	1.3	1.0	0.6	0.5
	212/218 ratio	1.3	0.8	1.2	1.0

6. *Construction, Purification, and Testing of 4-4-20/CC49 Heterodimer F_v*

The goals of this experiment were to produce, purify and analyze for activity a new heterodimer Fv that would bind to both fluorescein and the pan-carcinoma antigen TAG-72. The design consisted of two polypeptide chains, which associated to form the active heterodimer Fv. Each polypeptide chain can be described as a mixed single-chain Fv (mixed sFv). The first mixed sFv (GX 8952) comprised a 4-4-20 variable light chain (V_L) and a CC49 variable heavy chain (V_H) connected by a 217 polypeptide linker (Figure 3). The second mixed sFv (GX 8953) comprised a CC49 V_L and a 4-4-20 V_H connected by a 217 polypeptide linker (Figure 4). The sequence of the 217 polypeptide linker is shown in Table 2.

Results

A. *Purification*

One 10-liter fermentation of the *E. coli* production strain for each mixed sFv was grown on casein digest-glucose-salts medium at 32°C to an optical density at 600 nm of 15 to 20. The mixed sFv expression was induced by raising the temperature of the fermentation to 42°C for one hour. 277gm (wet cell weight) of *E. coli* GX 8952 and 233gm (wet cell weight) of *E. coli* GX 8953 were harvested in a centrifuge at 7000g for 10 minutes. The cell pellets were kept and the supernate discarded. The cell pellets were frozen at -20°C for storage.

2.55 liters of lysis/wash buffer (50mM Tris/ 200mM NaCl/ 1 mM EDTA, pH 8.0) was added to both of the mixed sFv's cell pellets, which were previously thawed and combined to give 510gm of total wet cell weight. After complete suspension of the cells they were then passed through a Gaulin homogenizer at 9000psi and 4°C. After this first pass the temperature

increased to 23°C. The temperature was immediately brought down to 0°C using dry ice and methanol. The cell suspension was passed through the Gaulin homogenizer a second time and centrifuged at 8000 rpm with a Dupont GS-3 rotor for 60 minutes. The supernatant was discarded after centrifugation and the pellets resuspended in 2.5 liters of lysis/wash buffer at 4°C. This suspension was centrifuged for 45 minutes at 8000 rpm with the Dupont GS-3 rotor. The supernatant was again discarded and the pellet weighed. The pellet weight was 136.1 gm.

1300ml of 6M Guanidine Hydrochloride/50mM Tris/50mM KCl/10mM $CaCl_2$ pH 8.0 at 4°C was added to the washed pellet. An overhead mixer was used to speed solubilization. After one hour of mixing, the heterodimer GuHCl extract was centrifuged for 45 minutes at 8000 rpm and the pellet was discarded. The 1425ml of heterodimer Fv 6M GuHCl extract was slowly added (16 ml/min) to 14.1 liters of Refold Buffer (50mM Tris/50mM KCl/10mM $CaCl_2$, pH 8.0) under constant mixing at 4°C to give an approximate dilution of 1:10. Refolding took place overnight at 4°C.

After 17 hours of refolding the anti-fluorescein activity was checked by a 40% quenching assay, and the amount of active protein calculated. 150mg total active heterodimer Fv was found by the 40% quench assay, assuming a 54,000 molecular weight.

4 liters of prechilled (4°C) 190 proof ethanol was added to the 15 liters of refolded heterodimer with mixing for 3 hours. The mixture sat overnight at 4°C. A flocculent precipitate had settled to the bottom after this overnight treatment. The nearly clear solution was filtered through a Millipak-200 (0.22μ) filter so as to not disturb the precipitate. A 40% quench assay showed that 10% of the anti-fluorescein activity was recovered in the filtrate.

The filtered sample of heterodimer was dialyzed, using a Pellicon system containing 10,000 dalton MWCO membranes, with dialysis buffer (40mM MOPS/0.5mM CaAcetate, pH 6.4) at 4°C. 20 liters of dialysis buffer was required before the conductivity of the retentate was equal to that of the dialysis buffer ($\sim$500μS). After dialysis the heterodimer sample was filtered

through a Millipak-20 filter, O.22μ. After this step a 40% quench assay showed there was 8.8 mg of active protein.

The crude heterodimer sample was loaded on a Poly CAT A cation exchange column at 20ml/min. The column was previously equilibrated with 60mM MOPS, 1 mM Calcium Acetate (CaAc) pH 6.4, at 4°C, (Buffer A). After loading, the column was washed with 150ml of Buffer A at 15ml/min. A 50min linear gradient was performed at 15ml/min using Buffer A and Buffer B (60mM MOPS, 20mM CaAc pH 7.5 at 4°C). The gradient conditions are presented in Table 5. Buffer C comprises 60mM MOPS, 100mM $CaCl_2$, pH 7.5.

Table 5

Time	%A	%B	%C	Flow
0:00	100.0	0.0	0.0	15ml/min
50:00	0.0	100.0	0.0	15ml/min
52:00	0.0	100.0	0.0	15ml/min
54:00	0.0	0.0	100.0	15ml/min
58:00	0.0	0.0	100.0	15ml/min
60:00	100.0	0.0	0.0	15ml/min

Approximately 50ml fractions were collected and analyzed for activity, purity, and molecular weight by size-exclusion chromatography. The fractions were not collected by peaks, so contamination between peaks is likely. Fractions 3 through 7 were pooled (total volume - 218ml), concentrated to 50ml and dialyzed against 4 liters of 60mM MOPS, 0.5mM CaAc pH 6.4 at 4°C overnight. The dialyzed pool was filtered through a O.22μ filter and checked for absorbance at 280nm. The filtrate was loaded onto the PolyCAT A column, equilibrated with 60mM MOPS, 1 mM CaAc pH 6.4 at 4°C, at a

flow rate of 10ml/min. Buffer B was changed to 60mM MOPS, 10mM CaAc pH 7.5 at 4°C. The gradient was run as in Table 5. The fractions were collected by peak and analyzed for activity, purity, and molecular weight. The chromatogram is shown in Figure 5. Fraction identification and analysis is presented in Table 6.

Table 6

Fraction Analysis of the Heterodimer Fv protein

Fraction No.	A_{280} reading	Total Volume (ml)	HPLE-SE Elution Time (min)
2	0.161	36	20.525
3	0.067	40	
4	0.033	40	
5	0.178	45	19.133
6	0.234	50	19.163
7	0.069	50	
8	0.055	40	

Fractions 2 to 7 and the starting material were analyzed by SDS gel electrophoresis, 4-20%. A picture and description of the gel is presented in Figure 6.

B. HPLC size exclusion results

Fractions 2, 5, and 6 correspond to the three main peaks in Figure 5 and therefore were chosen to be analyzed by HPLC size exclusion. Fraction 2 corresponds to the peak that runs at 21.775 minutes in the preparative purification (Figure 5), and runs on the HPLC sizing column at 20.525 minutes, which is in the monomeric position (Figure 7). Fractions 5 and 6

(30.1 and 33.455 minutes, respectively, in Figure 5) run on the HPLC sizing column (Figures 8 and 9) at 19.133 and 19.163 minutes, respectively (see Table 6). Therefore, both of these peaks could be considered dimers. 40% Quenching assays were performed on all fractions of this purification. Only fraction 5 gave significant activity. 2.4 mg of active CC49/4-4-20 heterodimer Fv was recovered in fraction 5, based on the Scatchard analysis described below.

C. N-terminal sequencing of the fractions

The active heterodimer Fv faction should contain both polypeptide chains. Internal sequence analysis showed that fractions 5 and 6 displayed N-terminal sequences consistent with the presence of both the Gx8952 and Gx8953 polypeptides and fraction 2 displayed a single sequence corresponding to the Gx8953 polypeptide only. We believe that fraction 6 was contaminated by fraction 5 (see Figure 5) since only fraction 5 had significant activity.

D. Anti-fluorescein activity by Scatchard analysis

The fluorescein association constants (Ka) were determined for fractions 5 and 6 using the fluorescence quenching assay described by Herron, J.N., in *Fluorescence Hapten: An Immunological Probe*, E.W. Voss, ed., CRC Press, Boca Raton, FL (1984). Each sample was diluted to approximately 5.0×10^{-8} M with 20 mM HEPES buffer pH 8.0. 590 μl of the 5.0×10^{-8} M sample was added to a cuvette in a fluorescence spectrophotometer equilibrated at room temperature. In a second cuvette 590 μl of 20 mM HEPES buffer pH 8.0 was added. To each cuvette was added 10 μl of 3.0×10^{-7} M fluorescein in 20 mM HEPES buffer pH 8.0, and the fluorescence recorded. This is repeated until 140 μl of fluorescein had been added. The resulting Scatchard analysis for fraction 5 shows a binding constant of 1.16×10^{9} M^{-1} for fraction #5 (see Figure 10). This is very close

to the 4-4-20/212 sFv constant of 1.1 x 10^9 M^{-1} (see Pantoliano *et al.*, *Biochemistry 30*:10117-10125 (1991)). The R intercept on the Scatchard analysis represents the fraction of active material. For fraction 5, 61% of the material was active. The graph of the Scatchard analysis on fraction 6 shows a binding constant of 3.3 x 10^8 M^{-1} and 14% active. The activity that is present in fraction 6 is most likely contaminants from fraction 5.

E. Anti-TAG-72 activity by competition ELISA

The CC49 monoclonal antibody was developed by Dr. Jeffrey Schlom's group, Laboratory of Tumor Immunology and Biology, National Cancer Institute. It binds specifically to the pan-carcinoma tumor antigen TAG-72. See Muraro, R., *et al.*, *Cancer Research 48*:4588-4596 (1988).

To determine the binding properties of the bivalent CC49/4-4-20 Fv (fraction 5) and the CC49/212 sFv, a competition enzyme-linked immunosorbent assay (ELISA) was set up in which a CC49 IgG labeled with biotin was competed against unlabeled CC49/4-4-20 Fv and the CC49/212 sFv for binding to TAG-72 on a human breast carcinoma extract (see Figure 11). The amount of biotin-labeled CC49 IgG was determined using avidin, biotin coupled to horse radish peroidase in a preformed complex and o-phenylene diamine dihydrochloride (OPD). The reaction was stopped after 10 min. with 4N sulfuric acid (H_2SO_4) and the optical density read at 490 nm. This competition ELISA showed that the bivalent CC49 4-4-20 Fv binds to the TAG-72 antigen. The CC49/4-4-20 Fv needed a two hundred-fold higher protein concentration to displace the IgG than the single-chain Fv.

Conclusions

We have produced a heterodimer Fv from two complementary mixed sFv's which has been shown to have the size of a dimer of the sFv's. The N-terminal analysis has shown that the active heterodimer Fv contains two

polypeptide chains. The heterodimer Fv has been shown to be active for both fluorescein and TAG-72 binding.

All references mentioned herein are incorporated by reference into this disclosure.

Having now fully described the invention by way of illustration and example for purposes of clarity and understanding, it will be apparent to those of ordinary skill in the art that certain changes and modifications may be practiced within the scope of the invention, as limited only by the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: **Enzon, Inc.**

(ii) TITLE OF INVENTION: Linker For Linked Fusion Polypeptides

(iii) NUMBER OF SEQUENCES: 14

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Sterne, Kessler, Goldstein & Fox
(B) STREET: 1100 New York Avenue, N.W.
(C) CITY: Washington
(D) STATE: D.C.
(E) COUNTRY: U.S.A.
(F) ZIP: 20005-3934

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: (to be assigned)
(B) FILING DATE: Herewith
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 07/980,529
(B) FILING DATE: 20-NOV-1992

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/002,845
(B) FILING DATE: 15-JAN-1993

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Goldstein, Jorge A.
(B) REGISTRATION NUMBER: 29,021
(C) REFERENCE/DOCKET NUMBER: 0977.2006604/JAG

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (202) 371-2600
(B) TELEFAX: (202) 371-2540

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: both

(ix) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION: 8
(D) OTHER INFORMATION: /label= Identification
/note= "The amino acid at position 8 is charged and a preferred embodiment of this amino acid is lysine or arginine."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Gly Ser Thr Ser Gly Ser Gly Xaa Pro Gly Ser Gly Glu Gly Ser Thr
1 5 10 15

Lys Gly

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: both

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Lys Glu Ser Gly Ser Val Ser Ser Glu Gln Leu Ala Gln Phe Arg Ser
1 5 10 15

Leu Asp

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: both

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Val Arg Gly Ser Pro Ala Ile Asn Val Ala Val His Val Phe
1 5 10

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: both

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Ala Gln Gly Thr Leu Ser Pro Ala Asp Lys Thr Asn Val Lys Ala Ala
1 5 10 15

Trp Gly Lys Val Met Thr
20

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: both

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Val Glu Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly

1 5 10 15

Val Asp

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: both

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Gly Ser Ala Ser Ala Pro Lys Leu Glu Glu Gly Glu Phe Ser Glu Ala
1 5 10 15

Arg Glu

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: both

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Gly Lys Ser Ser Gly Ser Gly Ser Glu Ser Lys Ser
1 5 10

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: both

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Gly Ser Thr Ser Gly Ser Gly Lys Ser Ser Glu Gly Lys Gly
1 5 10

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: both

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Gly Ser Thr Ser Gly Ser Gly Lys Ser Ser Glu Gly Ser Gly Ser Thr
1 5 10 15

Lys Gly

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: both

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Gly Ser Thr Ser Gly Ser Gly Lys Pro Gly Ser Gly Glu Gly Ser Thr
1               5                  10                  15

Lys Gly
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 725 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..723

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
GAC GTC GTT ATG ACT CAG ACA CCA CTA TCA CTT CCT GTT AGT CTA GGT      48
Asp Val Val Met Thr Gln Thr Pro Leu Ser Leu Pro Val Ser Leu Gly
  1               5                  10                  15

GAT CAA GCC TCC ATC TCT TGC AGA TCT AGT CAG AGC CTT GTA CAC AGT      96
Asp Gln Ala Ser Ile Ser Cys Arg Ser Ser Gln Ser Leu Val His Ser
            20                  25                  30

AAT GGA AAC ACC TAT TTA CGT TGG TAC CTG CAG AAG CCA GGC CAG TCT     144
Asn Gly Asn Thr Tyr Leu Arg Trp Tyr Leu Gln Lys Pro Gly Gln Ser
        35                  40                  45

CCA AAG GTC CTG ATC TAC AAA GTT TCC AAC CGA TTT TCT GGG GTC CCA     192
Pro Lys Val Leu Ile Tyr Lys Val Ser Asn Arg Phe Ser Gly Val Pro
    50                  55                  60

GAC AGG TTC AGT GGC AGT GGA TCA GGG ACA GAT TTC ACA CTC AAG ATC     240
Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Lys Ile
65                  70                  75                  80

AGC AGA GTG GAG GCT GAG GAT CTG GGA GTT TAT TTC TGC TCT CAA AGT     288
Ser Arg Val Glu Ala Glu Asp Leu Gly Val Tyr Phe Cys Ser Gln Ser
                85                  90                  95

ACA CAT GTT CCG TGG ACG TTC GGT GGA GGC ACC AAG CTT GAA ATC AAA     336
Thr His Val Pro Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Lys
            100                 105                 110

GGT TCT ACC TCT GGT AAA CCA TCT GAA GGC AAA GGT CAG GTT CAG CTG     384
Gly Ser Thr Ser Gly Lys Pro Ser Glu Gly Lys Gly Gln Val Gln Leu
        115                 120                 125
```

```
CAG CAG TCT GAC GCT GAG TTG GTG AAA CCT GGG GCT TCA GTG AAG ATT      432
Gln Gln Ser Asp Ala Glu Leu Val Lys Pro Gly Ala Ser Val Lys Ile
    130                 135                 140

TCC TGC AAG GCT TCT GGC TAC ACC TTC ACT GAC CAT GCA ATT CAC TGG      480
Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp His Ala Ile His Trp
145                 150                 155                 160

GTG AAA CAG AAC CCT GAA CAG GGC CTG GAA TGG ATT GGA TAT TTT TCT      528
Val Lys Gln Asn Pro Glu Gln Gly Leu Glu Trp Ile Gly Tyr Phe Ser
                165                 170                 175

CCC GGA AAT GAT GAT TTT AAA TAC AAT GAG AGG TTC AAG GGC AAG GCC      576
Pro Gly Asn Asp Asp Phe Lys Tyr Asn Glu Arg Phe Lys Gly Lys Ala
            180                 185                 190

ACA CTG ACT GCA GAC AAA TCC TCC AGC ACT GCC TAC GTG CAG CTC AAC      624
Thr Leu Thr Ala Asp Lys Ser Ser Ser Thr Ala Tyr Val Gln Leu Asn
        195                 200                 205

AGC CTG ACA TCT GAG GAT TCT GCA GTG TAT TTC TGT ACA AGA TCC CTG      672
Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys Thr Arg Ser Leu
    210                 215                 220

AAT ATG GCC TAC TGG GGT CAA GGA ACC TCA GTC ACC GTC TCC TAA TAG      720
Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val Thr Val Ser  *   *
225                 230                 235                 240

GAT CC                                                              725
Asp
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 241 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Asp Val Val Met Thr Gln Thr Pro Leu Ser Leu Pro Val Ser Leu Gly
 1               5                  10                  15

Asp Gln Ala Ser Ile Ser Cys Arg Ser Ser Gln Ser Leu Val His Ser
            20                  25                  30

Asn Gly Asn Thr Tyr Leu Arg Trp Tyr Leu Gln Lys Pro Gly Gln Ser
        35                  40                  45

Pro Lys Val Leu Ile Tyr Lys Val Ser Asn Arg Phe Ser Gly Val Pro
    50                  55                  60

Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Lys Ile
65                  70                  75                  80

Ser Arg Val Glu Ala Glu Asp Leu Gly Val Tyr Phe Cys Ser Gln Ser
                85                  90                  95

Thr His Val Pro Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Lys
            100                 105                 110
```

```
Gly Ser Thr Ser Gly Lys Pro Ser Glu Gly Lys Gly Gln Val Gln Leu
        115                 120                 125

Gln Gln Ser Asp Ala Glu Leu Val Lys Pro Gly Ala Ser Val Lys Ile
    130                 135                 140

Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp His Ala Ile His Trp
145                 150                 155                 160

Val Lys Gln Asn Pro Glu Gln Gly Leu Glu Trp Ile Gly Tyr Phe Ser
                165                 170                 175

Pro Gly Asn Asp Asp Phe Lys Tyr Asn Glu Arg Phe Lys Gly Lys Ala
            180                 185                 190

Thr Leu Thr Ala Asp Lys Ser Ser Ser Thr Ala Tyr Val Gln Leu Asn
        195                 200                 205

Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys Thr Arg Ser Leu
    210                 215                 220

Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val Thr Val Ser  *   *
225                 230                 235                 240

Asp
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 738 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..738

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
GAC GTC GTG ATG TCA CAG TCT CCA TCC TCC CTA CCT GTG TCA GTT GGC      48
Asp Val Val Met Ser Gln Ser Pro Ser Ser Leu Pro Val Ser Val Gly
  1               5                  10                  15

GAG AAG GTT ACT TTG AGC TGC AAG TCC AGT CAG AGC CTT TTA TAT AGT      96
Glu Lys Val Thr Leu Ser Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser
            20                  25                  30

GGT AAT CAA AAG AAC TAC TTG GCC TGG TAC CAG CAG AAA CCA GGG CAG     144
Gly Asn Gln Lys Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln
        35                  40                  45

TCT CCT AAA CTG CTG ATT TAC TGG GCA TCC GCT AGG GAA TCT GGG GTC     192
Ser Pro Lys Leu Leu Ile Tyr Trp Ala Ser Ala Arg Glu Ser Gly Val
    50                  55                  60

CCT GAT CGC TTC ACA GGC AGT GGA TCT GGG ACA GAT TTC ACT CTC TCC     240
Pro Asp Arg Phe Thr Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser
65                  70                  75                  80

ATC AGC AGT GTG AAG ACT GAA GAC CTG GCA GTT TAT TAC TGT CAG CAG     288
Ile Ser Ser Val Lys Thr Glu Asp Leu Ala Val Tyr Tyr Cys Gln Gln
```

```
                  85                  90                  95

TAT TAT AGC TAT CCC CTC ACG TTC GGT GCT GGG ACC AAG CTT GTG CTG      336
Tyr Tyr Ser Tyr Pro Leu Thr Phe Gly Ala Gly Thr Lys Leu Val Leu
            100                 105                 110

AAA GGC TCT ACT TCC GGT AAA CCA TCT GAA GGT AAA GGT GAA GTT AAA      384
Lys Gly Ser Thr Ser Gly Lys Pro Ser Glu Gly Lys Gly Glu Val Lys
        115                 120                 125

CTG GAT GAG ACT GGA GGA GGC TTG GTG CAA CCT GGG AGG CCC ATG AAA      432
Leu Asp Glu Thr Gly Gly Gly Leu Val Gln Pro Gly Arg Pro Met Lys
    130                 135                 140


CTC TCC TGT GTT GCC TCT GGA TTC ACT TTT AGT GAC TAC TGG ATG AAC      480
Leu Ser Cys Val Ala Ser Gly Phe Thr Phe Ser Asp Tyr Trp Met Asn
145                 150                 155                 160


TGG GTC CGC CAG TCT CCA GAG AAA GGA CTG GAG TGG GTA GCA CAA ATT      528
Trp Val Arg Gln Ser Pro Glu Lys Gly Leu Glu Trp Val Ala Gln Ile
                165                 170                 175

AGA AAC AAA CCT TAT AAT TAT GAA ACA TAT TAT TCA GAT TCT GTG AAA      576
Arg Asn Lys Pro Tyr Asn Tyr Glu Thr Tyr Tyr Ser Asp Ser Val Lys
            180                 185                 190

GGC AGA TTC ACC ATC TCA AGA GAT GAT TCC AAA AGT AGT GTC TAC CTG      624
Gly Arg Phe Thr Ile Ser Arg Asp Asp Ser Lys Ser Ser Val Tyr Leu
        195                 200                 205

CAA ATG AAC AAC TTA AGA GTT GAA GAC ATG GGT ATC TAT TAC TGT ACG      672
Gln Met Asn Asn Leu Arg Val Glu Asp Met Gly Ile Tyr Tyr Cys Thr
    210                 215                 220

GGT TCT TAC TAT GGT ATG GAC TAC TGG GGT CAA GGA ACC TCA GTC ACC      720
Gly Ser Tyr Tyr Gly Met Asp Tyr Trp Gly Gln Gly Thr Ser Val Thr
225                 230                 235                 240

GTC TCC TAA TAA GGA TCC                                             738
Val Ser  *   *  Gly Ser
                245
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 246 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Asp Val Val Met Ser Gln Ser Pro Ser Ser Leu Pro Val Ser Val Gly
  1               5                  10                  15

Glu Lys Val Thr Leu Ser Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser
            20                  25                  30

Gly Asn Gln Lys Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln
        35                  40                  45

Ser Pro Lys Leu Leu Ile Tyr Trp Ala Ser Ala Arg Glu Ser Gly Val
```

```
     50                  55                  60

Pro Asp Arg Phe Thr Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser
 65                  70                  75                  80

Ile Ser Ser Val Lys Thr Glu Asp Leu Ala Val Tyr Tyr Cys Gln Gln
                 85                  90                  95

Tyr Tyr Ser Tyr Pro Leu Thr Phe Gly Ala Gly Thr Lys Leu Val Leu
            100                 105                 110

Lys Gly Ser Thr Ser Gly Lys Pro Ser Glu Gly Lys Gly Glu Val Lys
        115                 120                 125

Leu Asp Glu Thr Gly Gly Gly Leu Val Gln Pro Gly Arg Pro Met Lys
    130                 135                 140

Leu Ser Cys Val Ala Ser Gly Phe Thr Phe Ser Asp Tyr Trp Met Asn
145                 150                 155                 160

Trp Val Arg Gln Ser Pro Glu Lys Gly Leu Glu Trp Val Ala Gln Ile
                165                 170                 175

Arg Asn Lys Pro Tyr Asn Tyr Glu Thr Tyr Tyr Ser Asp Ser Val Lys
            180                 185                 190

Gly Arg Phe Thr Ile Ser Arg Asp Asp Ser Lys Ser Ser Val Tyr Leu
        195                 200                 205

Gln Met Asn Asn Leu Arg Val Glu Asp Met Gly Ile Tyr Tyr Cys Thr
    210                 215                 220

Gly Ser Tyr Tyr Gly Met Asp Tyr Trp Gly Gln Gly Thr Ser Val Thr
225                 230                 235                 240

Val Ser  *   *  Gly Ser
                245
```

We claim:

1. A linked fusion polypeptide comprising a first polypeptide and a second polypeptide connected by a peptide linker, said peptide linker comprising one or more occurrences of the sequence XP, wherein X is a charged amino acid and said sequence is positioned within said peptide linker so as to inhibit proteolysis of said linked fusion polypeptide.

2. The linked fusion polypeptide of claim 1 wherein said first and second polypeptides are not derived from the same single chain protein or from the same chain of a multi-chain protein.

3. The linked fusion polypeptide of claim 2 wherein said first and second polypeptides are derived from different proteins.

4. The linked fusion polypeptide of claim 3 wherein said first and second polypeptides are derived from members of the immunoglobulin superfamily.

5. The linked fusion polypeptide of claim 4 wherein said first and second polypeptides are derived from immunoglobulins.

6. The linked fusion polypeptide of claim 5 wherein said linked fusion polypeptide is a mixed sFv.

7. The linked fusion polypeptide of claim 1 wherein said first and second polypeptides are derived from the same multi-chain protein.

8. The linked fusion polypeptide of claim 7 wherein said multi-chain protein is a member of the immunoglobulin superfamily.

9. The linked fusion polypeptide of claim 8 wherein said multi-chain protein is a T cell receptor.

10. The linked fusion polypeptide of claim 8 wherein said multi-chain protein is an immunoglobulin.

11. The fusion protein of claim 10 wherein said first polypeptide comprises the binding portion of the variable region of the heavy or light chain of said immunoglobulin.

12. The linked fusion polypeptide of claim 10 wherein said second polypeptide comprises the binding portion of the variable region of the heavy or light chain of said immunoglobulin.

13. The linked fusion polypeptide of claim 10 wherein said first polypeptide comprises the binding portion of the variable region of the heavy chain of said immunoglobulin and said second polypeptide comprises the binding portion of the variable region of the light chain of said immunoglobulin.

14. The linked fusion polypeptide of claim 1 wherein said peptide linker comprises about 10 to about 30 amino acids.

15. The linked fusion polypeptide of claim 14 wherein said peptide linker comprises at least 18 amino acids.

16. The linked fusion polypeptide of claim 15 wherein said sequence XP occurs at positions 8 and 9 from the amino terminus of said peptide linker.

17. The linked fusion polypeptide of claim 16 wherein said peptide linker comprises the amino acid sequence:

GSTSGSGXPGSGEGSTKG (SEQ ID No. 1).

18. The linked fusion polypeptide of claim 1 wherein said charged amino acid is a positively-charged amino acid.

19. The linked fusion polypeptide of claim 18 wherein said charged amino acid is lysine or arginine.

20. A DNA molecule coding for the linked fusion polypeptide of claim 1.

21. A peptide linker comprising a single amino acid chain of 18 to about 30 amino acids, said amino acid chain comprising the sequence:

GSTSGSGXPGSGEGSTKG (SEQ ID No. 1)

wherein X is a charged amino acid.

22. The peptide linker of claim 21 wherein said charged amino acid is a positively charged amino acid.

23. The peptide linker of claim 22 wherein said charged amino acid is lysine or arginine.

24. A DNA molecule coding for the peptide linker of claim 21.

25. A method of producing the linked fusion polypeptide of claim 1 in a host which comprises:

(a) providing a genetic sequence coding for said linked fusion polypeptide;

(b) transforming a host cell with said sequence;

(c) expressing said sequence in said host; and

(d) recovering said linked fusion polypeptide.

26. The method of claim 25 which further comprises purifying said linked fusion polypeptide after it is recovered.

27. The method of claim 25 wherein said host cell is a bacterial cell, yeast or other fungal cell, or a mammalian cell line.

28. The method of claim 25 wherein said linked fusion polypeptide is derived from one or more members of the immunoglobulin superfamily.

29. The method of claim 28 wherein said linked fusion polypeptide is derived from a T-cell receptor.

30. The method of claim 28 wherein said linked fusion polypeptide is derived from an immunoglobulin.

31. The method of claim 30 wherein said linked fusion polypeptide is an sFv.

32. The method of claim 28 wherein said linked fusion polypeptide is derived from two different immunoglobulins.

33. The method of claim 32 wherein said fusion protein is a mixed sFV.

34. A method of making a linked fusion polypeptide from a multi-chain protein, said method comprising:

(a) providing a first polypeptide corresponding to a first chain, or subfragment thereof, of said multi-chain protein;

(b) providing a second polypeptide corresponding to a second chain, or subfragment thereof, of said multi-chain protein;

(c) connecting said first polypeptide and said second polypeptide to opposite ends of a peptide linker to form said linked fusion polypeptide, said peptide linker comprising one or more occurrences of the sequence XP, wherein X is a charged amino acid and said sequence is positioned within said peptide linker so as to inhibit proteolysis of said linked fusion polypeptide; and

(d) recovering said linked fusion polypeptide.

35. The method of claim 34 wherein said multi-chain protein is a member of the immunoglobulin superfamily.

36. The method of claim 35 wherein said multi-chain protein is a T cell receptor.

37. The method of claim 35 wherein said multi-chain protein is an immunoglobulin.

38. The method of claim 37 wherein said first and second polypeptides comprise the binding portion of the variable region of the heavy or light chain of said immunoglobulin.

39. The method of claim 38 wherein said first polypeptide comprises the binding portion of the variable region of said immunoglobulin light chain and said second polypeptide comprises the binding portion of the variable region of said immunoglobulin heavy chain.

40. A method of making a linked fusion polypeptide from two different proteins, said method comprising:

(a) providing a first polypeptide corresponding to either a single chain protein or a chain of a multi-chain protein, or a subfragment thereof;

(b) providing a second polypeptide corresponding to either a single chain protein or a chain of a multi-chain protein different from that of said first polypeptide, or a subfragment thereof;

(c) connecting said first polypeptide and said second polypeptide to opposite ends of a peptide linker to form said linked fusion polypeptide, said peptide linker comprising one or more occurrences of the sequence XP, wherein X is a charged amino acid and said sequence is positioned within said peptide linker so as to inhibit proteolysis of said linked fusion polypeptide.

41. The method of claim 40 wherein said proteins are members of the immunoglobulin superfamily.

42. The method of claim 41 wherein said proteins are immunoglobulins.

43. The method of claim 42 wherein said first and second polypeptides comprise the binding portion of the variable region of the heavy or light chain of said immunoglobulins.

44. The method of claim 43 wherein said linked fusion polypeptide is a mixed sFV.

45. The linked fusion polypeptide of claim 1 wherein said first polypeptide is CC49 V_L, said second polypeptide is CC49 V_H, and said peptide linker comprises the amino acid sequence: GSTSGSGKPGSGEGSTKG (SEQ ID No. 10).

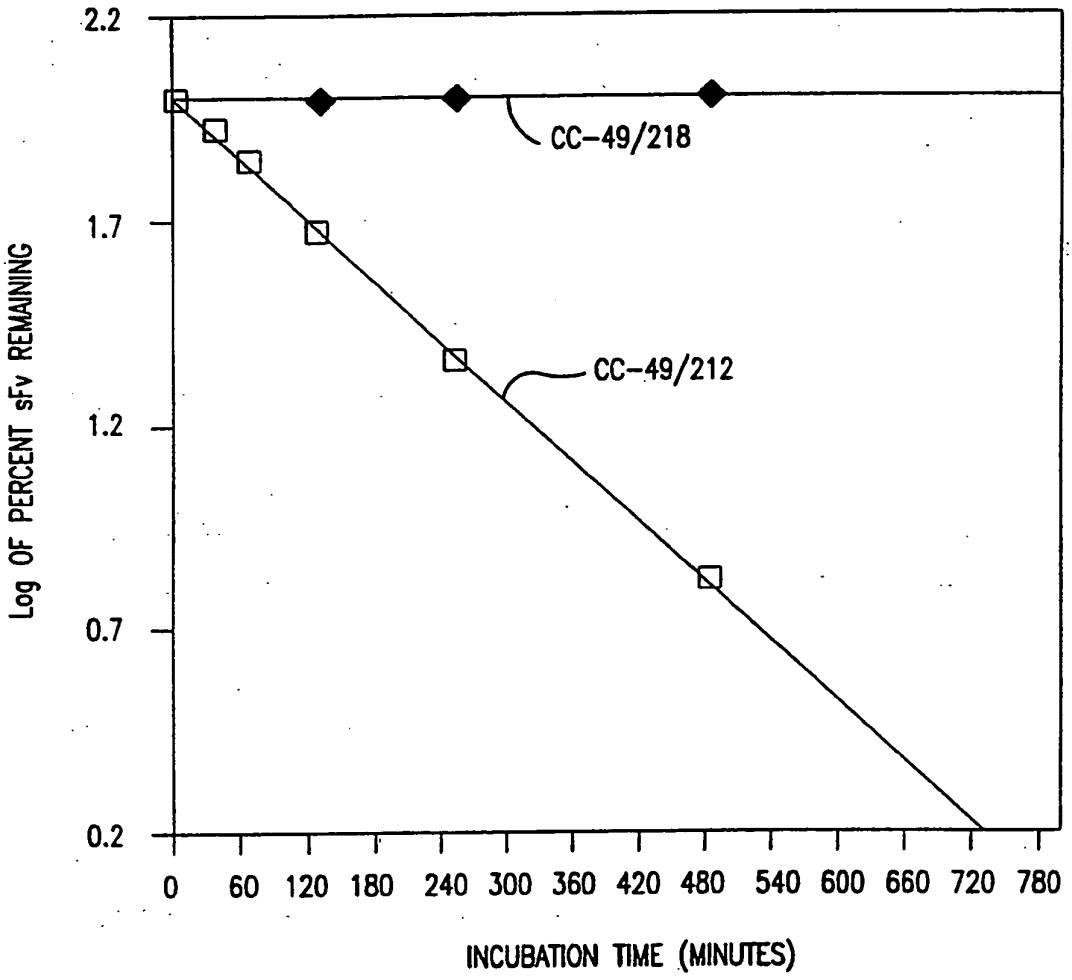

2.2
1.7
1.2
0.7
0.2
Log OF PERCENT sFv REMAINING
CC-49/218
CC-49/212
0
60
120
180
240
300
360
420
480
540
600
660
720
780
INCUBATION TIME (MINUTES)

FIG.1A

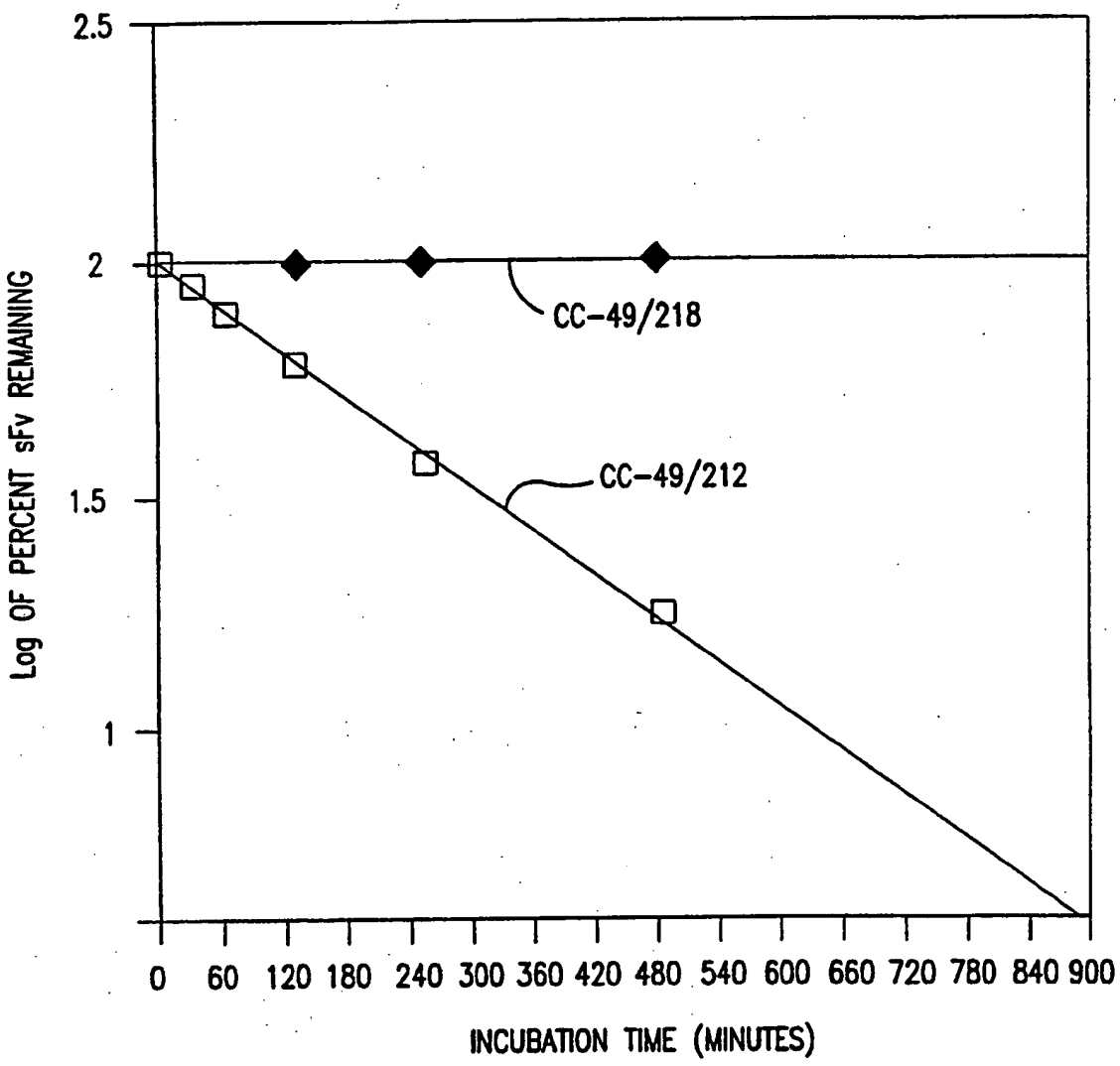

2.5
2
1.5
1
Log OF PERCENT sFv REMAINING
CC-49/218
CC-49/212
0 60 120 180 240 300 360 420 480 540 600 660 720 780 840 900
INCUBATION TIME (MINUTES)

FIG.1B

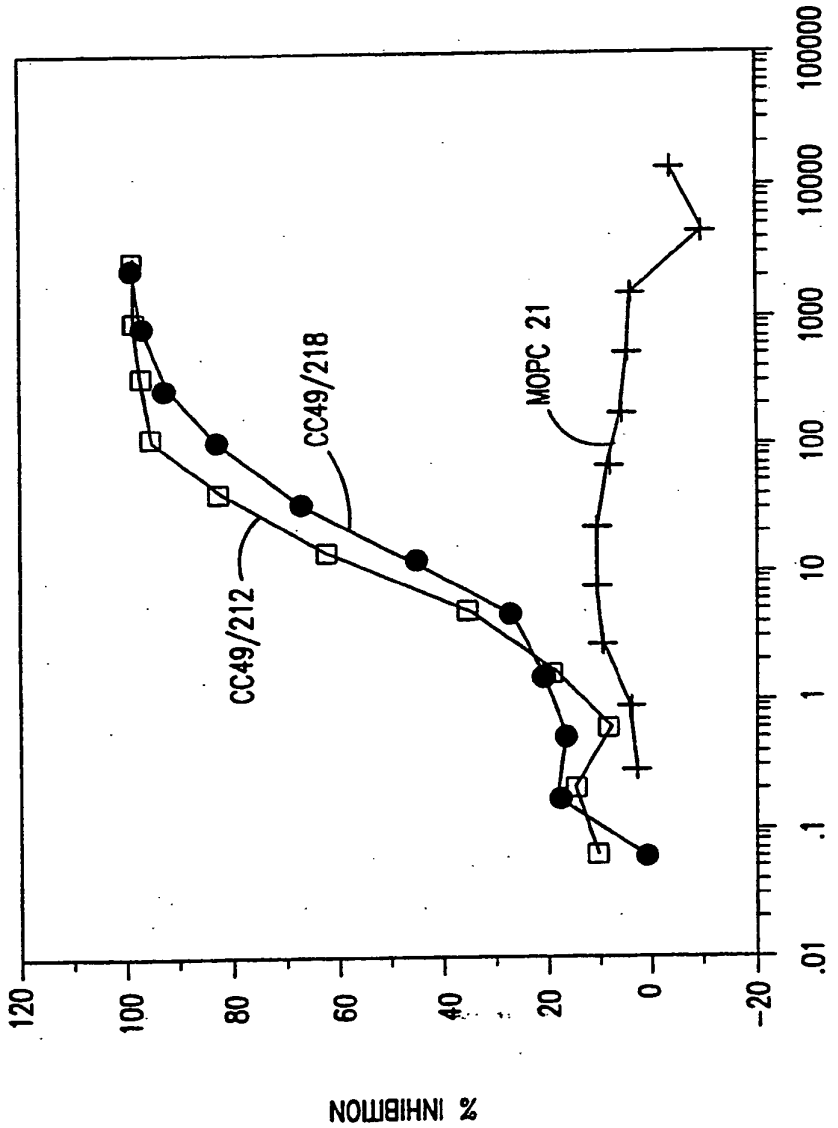

% INHIBITION
120
100
80
60
40
20
0
-20
.01
.1
1
10
100
1000
10000
100000
CONCENTRATION (nM)
CC49/212
CC49/218
MOPC 21

FIG.2

4-4-20 V_L / 217 / CC49 V_H gene

4-4-20 V_L 10 20
Asp Val Val Met Thr Gln Thr Pro Leu Ser Leu Pro Val Ser Leu Gly Asp Gln Ala Ser
GAC GTC GTT ATG ACT CAG ACA CCA CTA TCA CTT CCT GTT AGT CTA GGT GAT CAA GCC TCC
Aat II

30 40
Ile Ser Cys Arg Ser Ser Gln Ser Leu Val His Ser Asn Gly Asn Thr Tyr Leu Arg Trp
ATC TCT TGC AGA TCT AGT CAG AGC CTT GTA CAC AGT AAT GGA AAC ACC TAT TTA CGT TGG

50 60
Tyr Leu Gln Lys Pro Gly Gln Ser Pro Lys Val Leu Ile Tyr Lys Val Ser Asn Arg Phe
TAC CTG CAG AAG CCA GGC CAG TCT CCA AAG GTC CTG ATC TAC AAA GTT TCC AAC CGA TTT

70 80
Ser Gly Val Pro Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Lys Ile
TCT GGG GTC CCA GAC AGG TTC AGT GGC AGT GGA TCA GGG ACA GAT TTC ACA CTC AAG ATC

90 100
Ser Arg Val Glu Ala Glu Asp Leu Gly Val Tyr Phe Cys Ser Gln Ser Thr His Val Pro
AGC AGA GTG GAG GCT GAG GAT CTG GGA GTT TAT TTC TGC TCT CAA AGT ACA CAT GTT CCG

110 217 Linker 120
Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Lys Gly Ser Thr Ser Gly Lys Pro Ser
TGG ACG TTC GGT GGA GGC ACC AAG CTT GAA ATC AAA GGT TCT ACC TCT GGT AAA CCA TCT
Hind III

CC49 V_H 130 140
Glu Gly Lys Gly Gln Val Gln Leu Gln Gln Ser Asp Ala Glu Leu Val Lys Pro Gly Ala
GAA GGC AAA GGT CAG GTT CAG CTG CAG CAG TCT GAC GCT GAG TTG GTG AAA CCT GGG GCT
Pvu II PstI

150 160
Ser Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp His Ala Ile His Trp
TCA GTG AAG ATT TCC TGC AAG GCT TCT GGC TAC ACC TTC ACT GAC CAT GCA ATT CAC TGG

170 180
Val Lys Gln Asn Pro Glu Gln Gly Leu Glu Trp Ile Gly Tyr Phe Ser Pro Gly Asn Asp
GTG AAA CAG AAC CCT GAA CAG GGC CTG GAA TGG ATT GGA TAT TTT TCT CCC GGA AAT GAT

190 200
Asp Phe Lys Tyr Asn Glu Arg Phe Lys Gly Lys Ala Thr Leu Thr Ala Asp Lys Ser Ser
GAT TTT AAA TAC AAT GAG AGG TTC AAG GGC AAG GCC ACA CTG ACT GCA GAC AAA TCC TCC

210 220
Ser Thr Ala Tyr Val Gln Leu Asn Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys
AGC ACT GCC TAC GTG CAG CTC AAC AGC CTG ACA TCT GAG GAT TCT GCA GTG TAT TTC TGT

230 240
Thr Arg Ser Leu Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val Thr Val Ser *** ***
ACA AGA TCC CTG AAT ATG GCC TAC TGG GGT CAA GGA ACC TCA GTC ACC GTC TCC TAA TAG

Asp
GAT CC
Bam HI

FIG.3

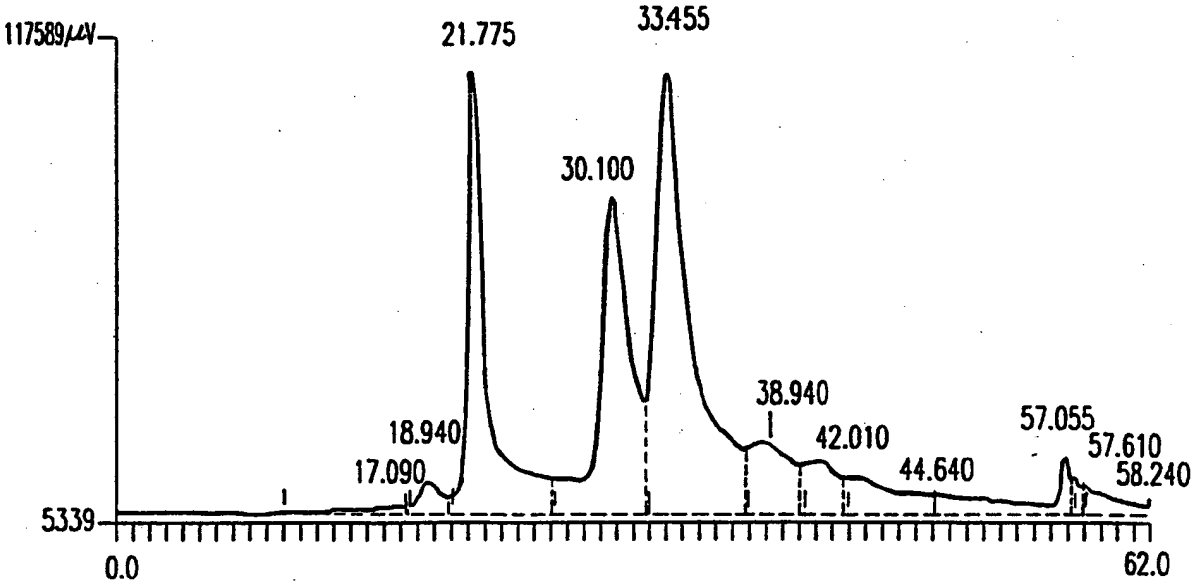
CC49 V_L / 217 / 4-4-20 V_H gene 5/12

CC49 V_L

```
                                    10                                  20
Asp Val Val Met Ser Gln Ser Pro Ser Ser Leu Pro Val Ser Val Gly Glu Lys Val Thr
GAC GTC GTG ATG TCA CAG TCT CCA TCC TCC CTA CCT GTG TCA GTT GGC GAG AAG GTT ACT
Aat II
                                    30                                  40
Leu Ser Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser Gly Asn Gln Lys Asn Tyr Leu Ala
TTG AGC TGC AAG TCC AGT CAG AGC CTT TTA TAT AGT GGT AAT CAA AAG AAC TAC TTG GCC
                                    50                                  60
Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Lys Leu Leu Ile Tyr Trp Ala Ser Ala Arg
TGG TAC CAG CAG AAA CCA GGG CAG TCT CCT AAA CTG CTG ATT TAC TGG GCA TCC GCT AGG
                                    70                                  80
Glu Ser Gly Val Pro Asp Arg Phe Thr Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser
GAA TCT GGG GTC CCT GAT CGC TTC ACA GGC AGT GGA TCT GGG ACA GAT TTC ACT CTC TCC
                                    90                                  100
Ile Ser Ser Val Lys Thr Glu Asp Leu Ala Val Tyr Tyr Cys Gln Gln Tyr Tyr Ser Tyr
ATC AGC AGT GTG AAG ACT GAA GAC CTG GCA GTT TAT TAC TGT CAG CAG TAT TAT AGC TAT
                                    110          217 Linker             120
Pro Leu Thr Phe Gly Ala Gly Thr Lys Leu Val Leu Lys Gly Ser Thr Ser Gly Lys Pro
CCC CTC ACG TTC GGT GCT GGG ACC AAG CTT GTG CTG AAA GGC TCT ACT TCC GGT AAA CCA
                                    Hind III
                        4-4-20 VH   140
Ser Glu Gly Lys Gly Glu Val Lys Leu Asp Glu Thr Gly Gly Gly Leu Val Gln Pro Gly
TCT GAA GGT AAA GGT GAA GTT AAA CTG GAT GAG ACT GGA GGA GGC TTG GTG CAA CCT GGG
                                    150                                 160
Arg Pro Met Lys Leu Ser Cys Val Ala Ser Gly Phe Thr Phe Ser Asp Tyr Trp Met Asn
AGG CCC ATG AAA CTC TCC TGT GTT GCC TCT GGA TTC ACT TTT AGT GAC TAC TGG ATG AAC
                                    170                                 180
Trp Val Arg Gln Ser Pro Glu Lys Gly Leu Glu Trp Val Ala Gln Ile Arg Asn Lys Pro
TGG GTC CGC CAG TCT CCA GAG AAA GGA CTG GAG TGG GTA GCA CAA ATT AGA AAC AAA CCT
                                    190                                 200
Tyr Asn Tyr Glu Thr Tyr Tyr Ser Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp
TAT AAT TAT GAA ACA TAT TAT TCA GAT TCT GTG AAA GGC AGA TTC ACC ATC TCA AGA GAT
                                    210                                 220
Asp Ser Lys Ser Ser Val Tyr Leu Gln Met Asn Asn Leu Arg Val Glu Asp Met Gly Ile
GAT TCC AAA AGT AGT GTC TAC CTG CAA ATG AAC AAC TTA AGA GTT GAA GAC ATG GGT ATC
                                    230                                 240
Tyr Tyr Cys Thr Gly Ser Tyr Tyr Gly Met Asp Tyr Trp Gly Gln Gly Thr Ser Val Thr
TAT TAC TGT ACG GGT TCT TAC TAT GGT ATG GAC TAC TGG GGT CAA GGA ACC TCA GTC ACC

Val Ser *** *** Gly Ser
GTC TCC TAA TAA GGA TCC
                Bam HI
```

FIG.4



Analysis: Channel A

Peak No.	Time	Type	Height(μV)	Area(μV-sec)	Area%
1	17.090	N1	1651	348239	0.778
2	18.940	N2	8014	669441	1.496
3	21.775	N3	104401	8617252	19.263
4	30.100	N4	74925	9753616	21.804
5	33.455	N5	106864	15749605	35.208
6	38.940	N6	17296	2833701	6.334
7	42.010	N7	12645	1637917	3.661
8	44.640	N8	9287	1968584	4.400
9	57.055	N9	13767	2012338	4.498
10	57.610	N10	9323	210914	0.471
11	58.240	X11	6824	930855	2.080
Total Area				44732462	99.993

FIG.5

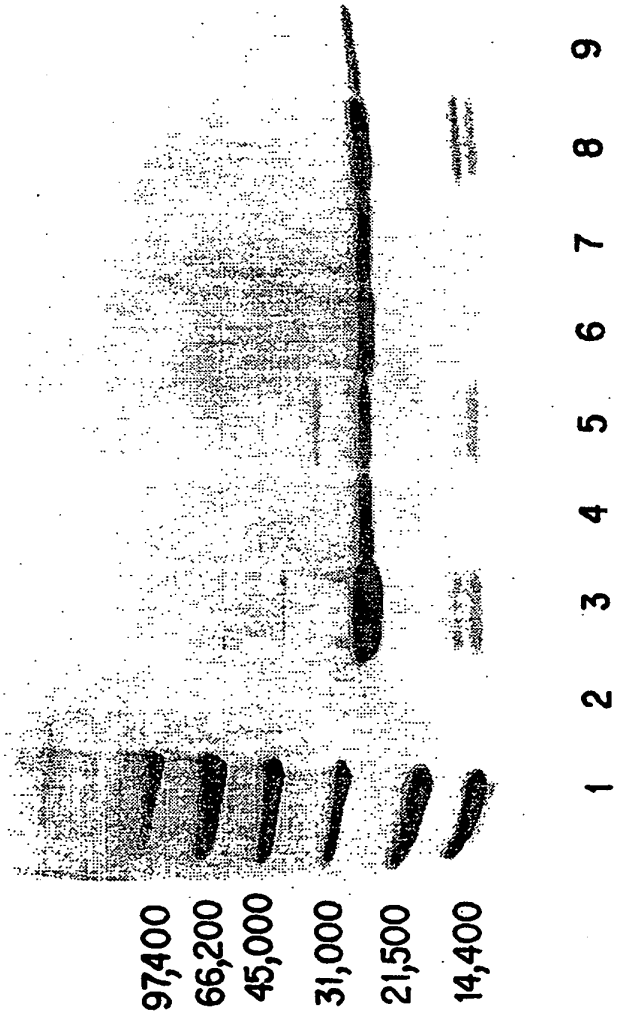


FIG.6

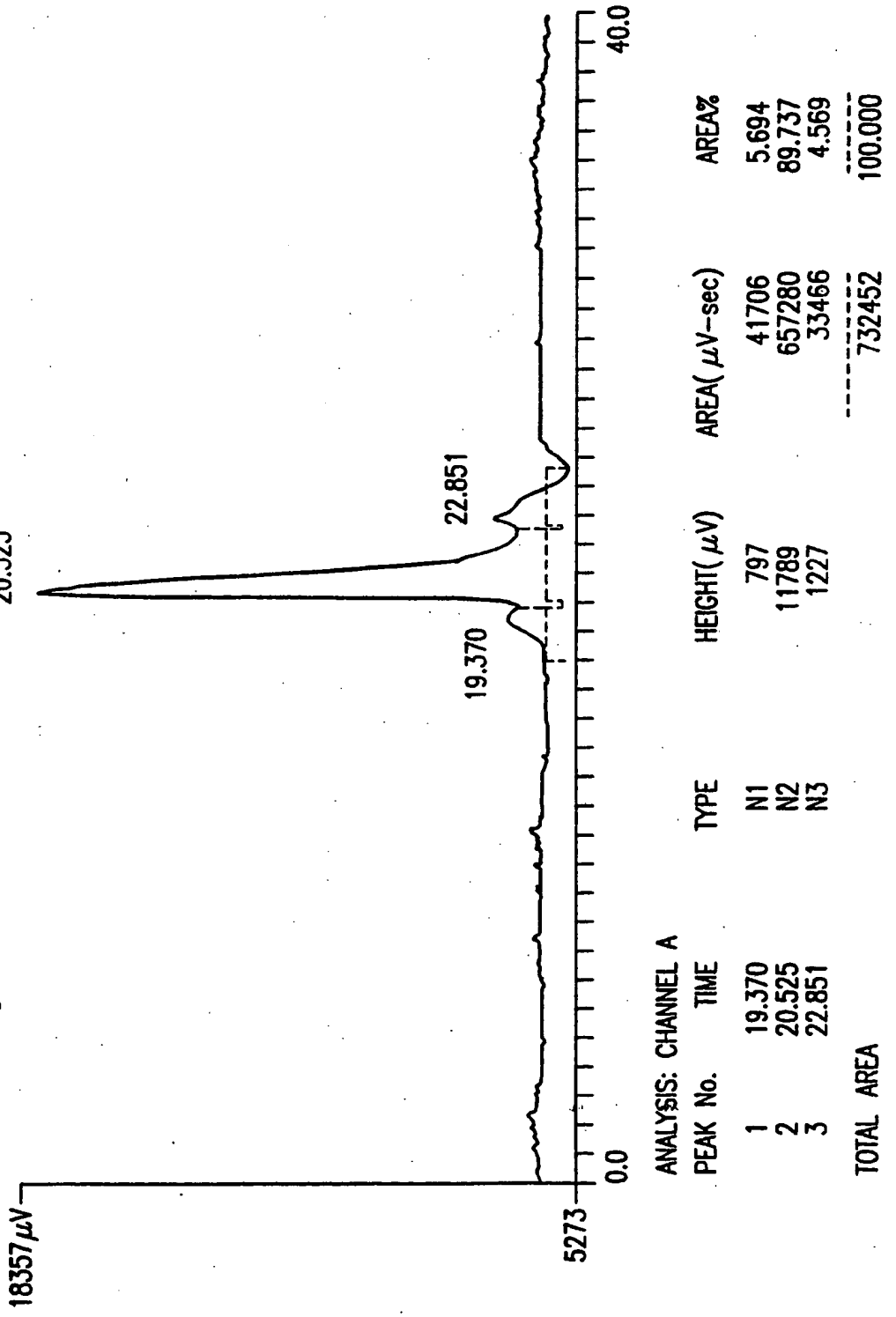

18357 μV
5273
0.0
40.0
20.525
22.851
19.370
ANALYSIS: CHANNEL A
PEAK No. TIME TYPE HEIGHT(μV) AREA(μV-sec) AREA%
1 19.370 N1 797 41706 5.694
2 20.525 N2 11789 657280 89.737
3 22.851 N3 1227 33466 4.569
TOTAL AREA 732452 100.000

FIG.7

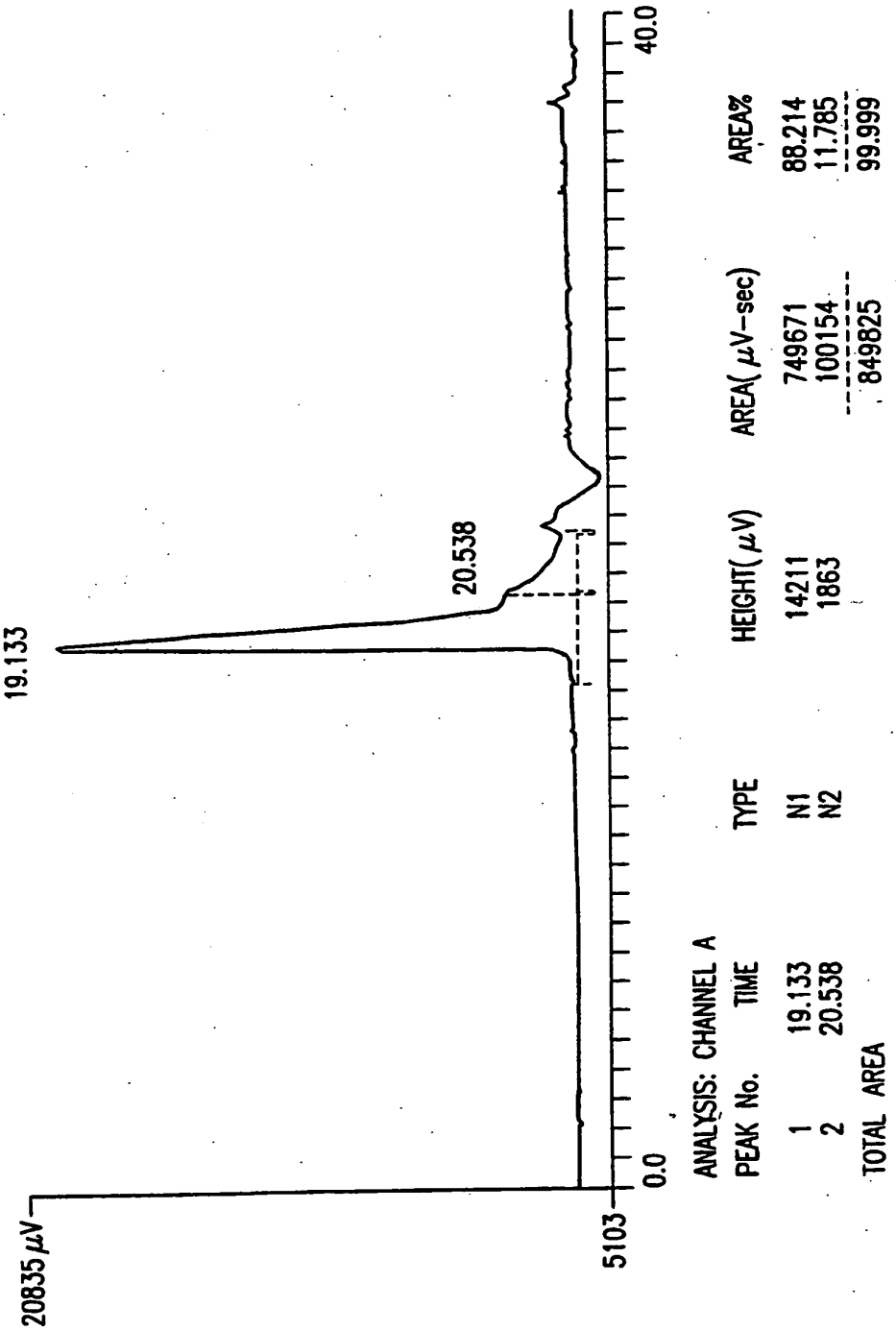

20835 μV
5103
0.0
40.0
19.133
20.538
ANALYSIS: CHANNEL A
PEAK No. TIME TYPE HEIGHT(μV) AREA(μV-sec) AREA%
1 19.133 N1 14211 749671 88.214
2 20.538 N2 1863 100154 11.785
TOTAL AREA 849825 99.999

FIG.8

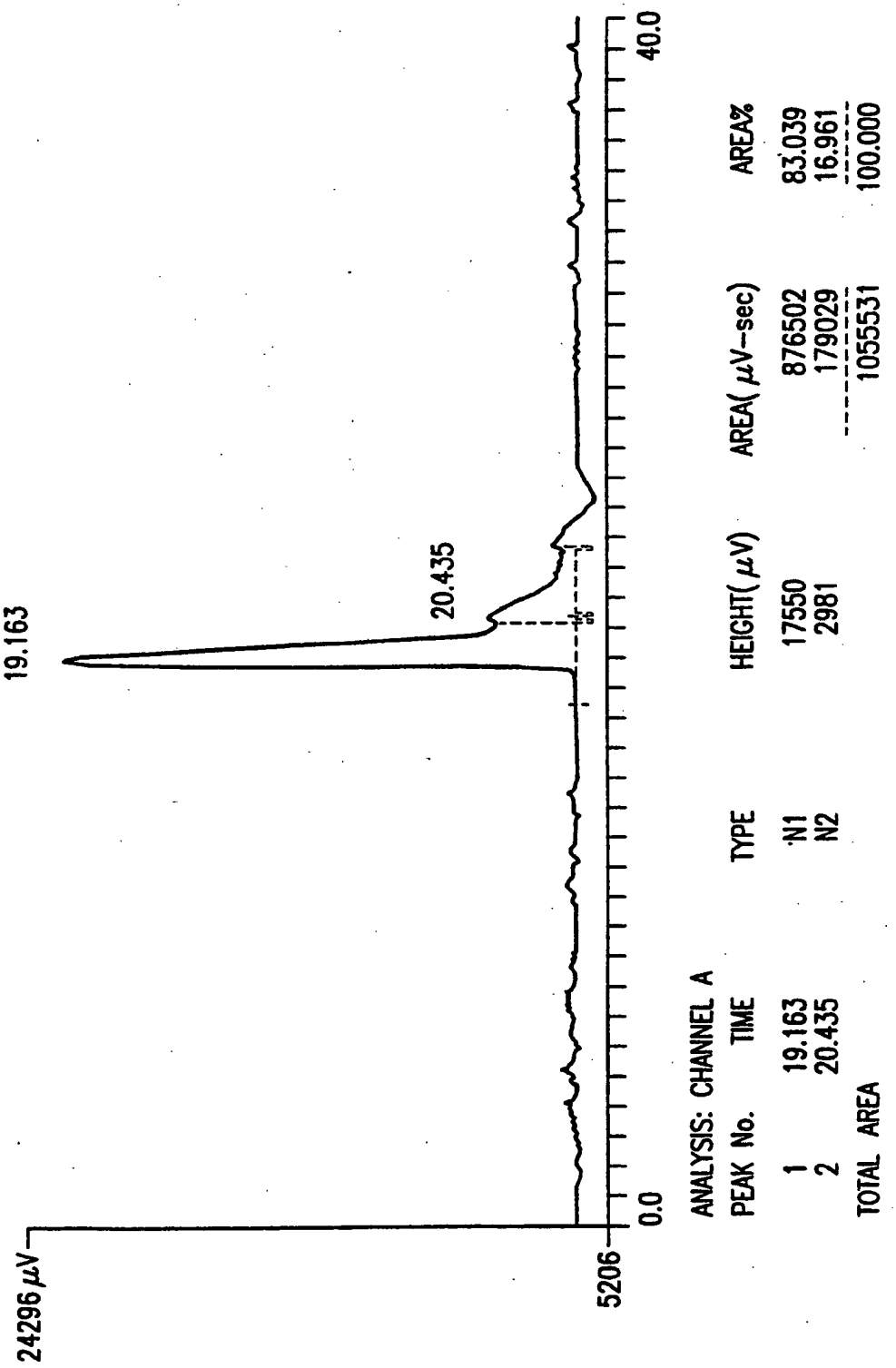

24296 μV
5206
0.0
40.0
19.163
20.435
ANALYSIS: CHANNEL A
PEAK No. TIME TYPE HEIGHT(μV) AREA(μV-sec) AREA%
1 19.163 N1 17550 876502 83.039
2 20.435 N2 2981 179029 16.961
TOTAL AREA 1055531 100.000

FIG.9

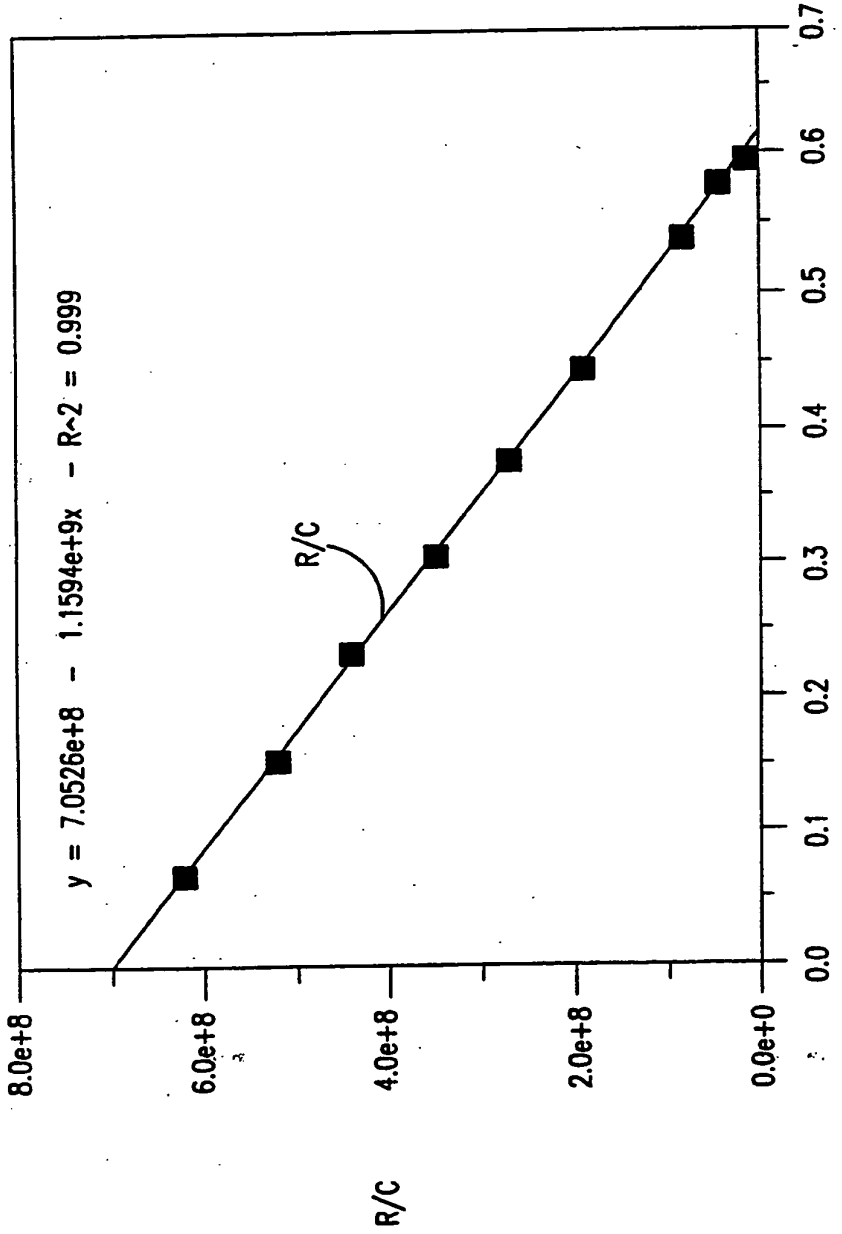

y = 7.0526e+8 - 1.1594e+9x - R^2 = 0.999
R/C
8.0e+8
6.0e+8
4.0e+8
2.0e+8
0.0e+0
R/C
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
R

FIG.10

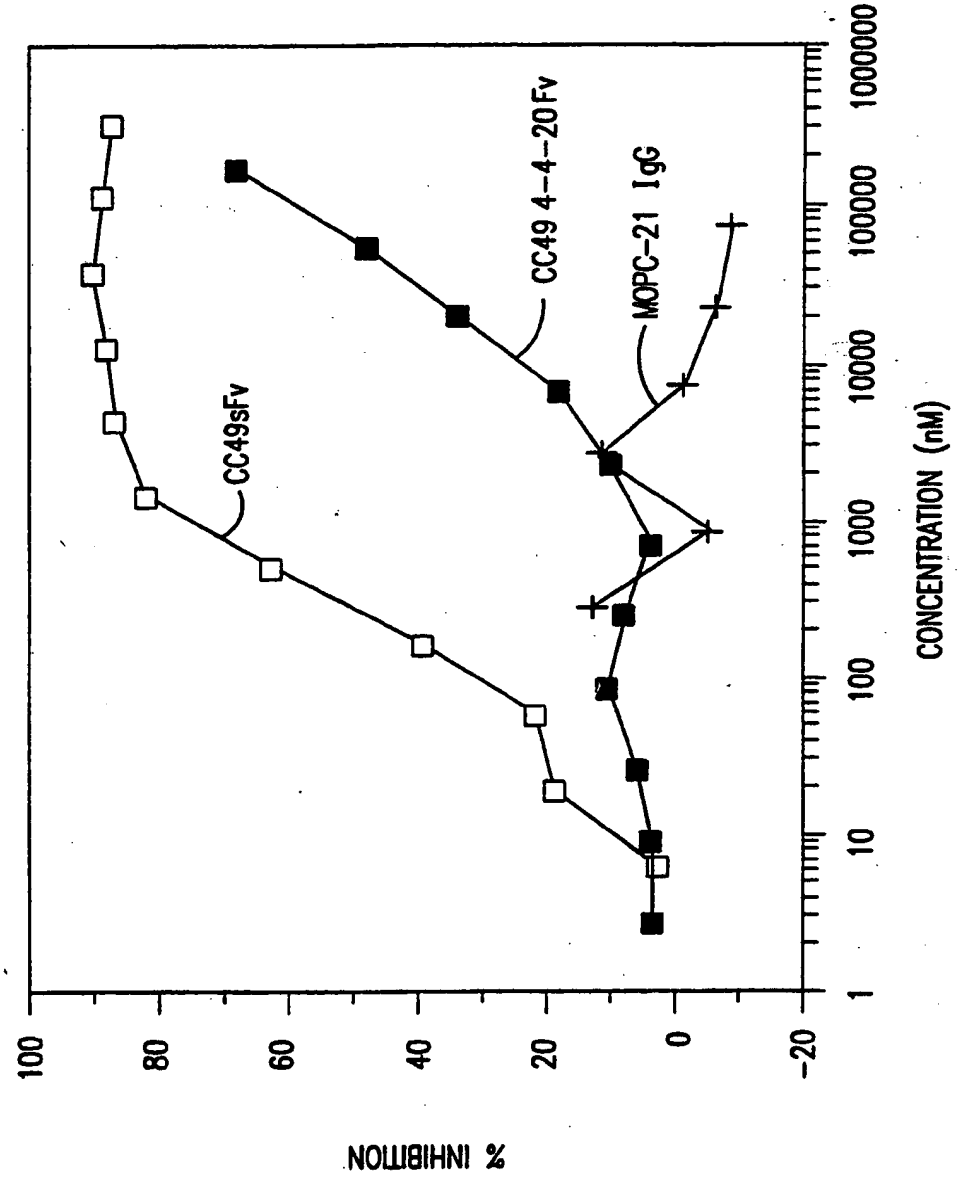

CC49sFv
CC49 4-4-20Fv
MOPC-21 IgG
% INHIBITION
CONCENTRATION (nM)
100
80
60
40
20
0
-20
1
10
100
1000
10000
100000
1000000

FIG.11